X-731-67-7

NASA TM X- 55747

CONFERENCE ON ADAPTIVE TELEMETRY HELD FEBRUARY 15-16, 1966

GPO PRICE $ _____

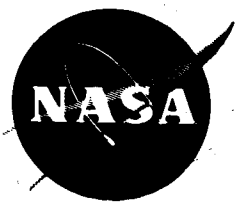
CFSTI PRICE(S) $ _____

Hard copy (HC) 3.00

Microfiche (MF) .65

JANUARY 1967

ff 653 July 65



GODDARD SPACE FLIGHT CENTER

GREENBELT, MARYLAND

X-731-67-7

CONFERENCE ON ADAPTIVE TELEMETRY

HELD

FEBRUARY 15-16, 1966

January 1967

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland

FOREWORD

The great potential of adaptive communications to all areas of the National Space Program, the growing need for more economic data transmission systems, and the lack of any previous unified action to bring developers and users together led to the Conference on Adaptive Telemetry, sponsored by NASA and held at Goddard Space Flight Center on February 15 and 16, 1966. This document contains those conference papers which were available at the time of printing with abstracts of unavailable papers appended in the interest of completeness.

The Conference proceedings were arranged into four sessions:

I Adaptive Prediction and Data Compression

II Adaptive Data Control and Processing

III Adaptive Encoding, Decoding, and Modulation

IV Adaptive Telemetry — Problems and Limitations

There was no attempt to compartmentalize the sessions precisely and restrict a topic — Data Compression, for instance — to a single session. Also, unfortunately, it has not been possible to capture the additional information and interpretations which came out in the discussion periods at the end of each conference session.

This conference was the first, to our knowledge, devoted solely to the subject of adaptive telemetry. Although the subject is relatively new and is being investigated by a relatively small number of researchers, the conference was enthusiastically supported and attended and offered a unique opportunity for the clarification of concepts and crossfertilization of ideas. Obviously, the work reported in this field had been done prior to the conference, but results had not been given wide circulation or close correlation.

From the standpoint of the quality of the papers, the wide representation of the participants, and the interest stimulated, the conference was very rewarding. The more significant results were (1) an exchange of views on the fundamental concepts of adaptive telemetry, (2) a better delineation of the various techniques encompassed by the term "adaptive," and (3) the identification of adaptive features in existing and proposed hardware implementations. The papers presented were fairly evenly divided among these objectives and the theme of the conference proceeded generally in that order. It is hoped that interest and work in adaptive telemetry will grow and important advances in the fields of

theory, techniques, and applications will be made. In particular, it is hoped that this conference has stimulated the interest of other potential users with whom we wish to open and maintain a close communication.

The original idea for a conference which would bring together together the proponents of adaptive telemetry was generated by Dr. Balakrishnan and his associates at UCLA about a year ago. A proposal to hold this conference at Goddard Space Flight Center received the immediate and enthusiastic support of Dr. John F. Clark, Director of the Center. Following the decision to hold the conference, strong support and valuable guidance was received from Dr. A. V. Balakrishnan at UCLA, Dr. L. D. Davisson of Princeton, Dr. J. C. Hancock of Purdue, Dr. J. Weber of USC, and researchers at Stanford University and SRI. From private industry, help was received from Dr. R. W. Sanders of Space General, Mr. L. Gardenhire of Radiation Inc., and research personnel of Astro-Power, EMR, and Hughes.

The number of those who participated in the conference, either as attendees, registrants, or in a more active capacity, exceeded 125 persons. Twenty-six papers were presented orally. The breadth of the participation in the conference can be seen by noting the affiliations of the authors. Representatives of six private companies presented papers. Five papers came from universities, six from the Jet Propulsion Laboratory, and nine from the Goddard Space Flight Center.

It is a great pleasure to acknowledge the support and contributions of the many people who helped to plan and conduct the conference, presented papers, and prepared these Proceedings. Particular credit should go to Dr. Balakrishnan who was cochairman of the conference, to Docters Benn Martin, George Ludwig, and Robert Rochelle who chaired separate sessions, and to Mr. Alfred Shehab of the Goddard Educational and Special Program Office and Mr. Charles Laughlin of the Systems Division who helped with the organization of the conference.

R. A. Stampfl
Greenbelt, Maryland
July, 1966

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

OPENING SESSION

Welcome — John F. Clark
Director, GSFC

Opening Remarks — A. V. Balakrishnan
Co-Chairman of Conference
UCLA

R. A. Stampfl
Co-Chairman of Conference
GSFC

WELCOME

CONFERENCE ON ADAPTIVE TELEMETRY

I regret that I will be unable to join you in what appears to be an interesting and timely meeting. However, I have reviewed with Dr. Stampfl your agenda for this two-day conference and I am pleased to note your interest and concern with the most effective management of large quantities of scientific data — an area of great importance to Goddard Space Flight Center.

Goddard is certainly a very logical place for this conference. Because of our assigned missions, we were among the first to be faced with these data management problems. Over the past 7 years, NASA has succeeded in putting into orbit dozens of satellites, both manned and unmanned, in order to learn more about our earth's natural environment so that we can apply the knowledge gained for the benefit of all mankind. This effort has involved the development of launch vehicles, spacecraft, and tracking and data systems. Literally hundreds of different instruments have been employed in space to acquire these data and to return them to Earth. The task of eliminating unnecessary redundancy from these data before transmission and their collection and processing for use by the scientific investigators are of vital concern to all of us involved in the management of these important space flight missions. Today, since we are confronted with the necessity of handling efficiently the constantly increasing data rates with essentially fixed resources of personnel and funds, the most competent resolution of the difficult problems involved is of the utmost importance.

I am therefore delighted that this conference will give the Goddard staff an opportunity to be exposed to your ideas and thinking, while you gain further insight into our needs and problems. Please accept my best wishes for a pleasant and fruitful meeting.

John F. Clark
Director, GSFC
Greenbelt, Maryland

OPENING REMARKS

SPACE EXPLORATION'S CHALLENGE

TO ADAPTIVE TELEMETRY*

Although the National Space Program has not suffered significantly to date from a limited satellite data transmission capability, there is hardly any area of the future program that cannot benefit from the application of adaptive telemetry. Many of those attending this conference are not only contributors to the discipline but also potential users. The most difficult problem for the user is the rigorous definition of the information which is essential to him. Once this goal is attained, vast quantities of data may not be needed, results of research will be available more quickly, and the search can proceed to the next phase. An adaptive system, by implication, is useful for many different sources of information and the constraints on, or assumptions about, the source are few. This is a vital characteristic from the users point of view because the exact nature of his data source is unknown. It is here where our enthusiasm for the art should be tempered with restraint, for the limits of adaptive telemetry may cause rejection of the entirely unexpected, and thus deprive us of discovery of the unanticipated.

GENERAL CONSIDERATIONS

Telemetry systems need not always be designed on a worst case basis. If a worst case situation can be presumed to be a transitory situation, then advantage should be taken of the intervals during which less severe requirements are placed on the system. In addition, telemetry systems are often called upon to handle highly redundant information sources. In such a case a telemetry system designer is obligated by the demands of other sources to take maximum advantage of this redundancy. Often, the ultimate usefulness of telemetered data is time and level dependent. This is definitely a nonquantitative aspect which transcends all theoretical information concepts; none the less, it is highly important. It is not difficult to provide examples where the transmission of unnecessary data (due to the inability of a system to adapt to changing conditions or requirements) seriously limits the overall effectiveness of a space mission. The more subtle advantages of adaptive techniques apply to virtually every future space program.

* A consolidation of the opening remarks of Dr. Balakrishnan, UCLA, and Dr. Stampfl, GSFC, chairman of the Conference on Adaptive Telemetry.

The need for transmitting only the most meaningful information is dramatized by the present and foreseeable telemetry limitations at planetary distances. With reasonable ground facilities, an 8-foot antenna reflector would permit less than 50 bits per second to be transmitted to Earth from a spacecraft at 5 astronomical units distance. Even with a 30-foot antenna reflector, which can reasonably be expected for space flight, this figure would only increase to a few thousand bits per second. Since such a telemetry system cannot be allowed to be choked up with redundant data, some form of adaptivity will be essential.

Closer to our planet, the time is already here when literally unbounded quantities of data can be collected continuously from satellites at geosynchronous altitudes. Most of Earth will soon be under continuous observation so that sensors not only can detect a multiplicity of parameters with high angular resolution but also can collect and distribute other kinds of data.

Optical sensors, like television cameras or infrared scanners, are usually quoted as data source examples because of their large bandwidth requirements. They can also be cited as the outstanding example of the disparity between the number of bits to be transmitted and the actual useful information content in the delivered image. As is well known, for weather observation only gross cloud patterns are fed into the prediction process. So far, it is only natural that adaptive compression techniques be applied to these data sources.

Improvement in weather predictions will depend upon progress in adaptive telemetry when large numbers of atmospheric soundings are to be performed continuously. The boundary conditions derived from these soundings are needed to update solutions of the atmospheric prediction equations. The economic benefits of accurate weather predictions that are valid for many days or perhaps weeks in advance are obvious.

As another example, the air traffic control problems in the next decade can be mentioned. The solutions to these problems will require more than just highly sophisticated versions of present technology. This is due largely to the characteristics of the supersonic transports currently under development. To be economically justifiable, these aircraft will have to be airborne virtually all of the time; thus, ground maintenance and turnaround times will be of paramount importance. Therefore, telemetry during flight will be needed to diagnose aircraft for servicing; also, many additional data will have to be collected for central processing. Telemetry systems aboard the supersonic transport will have much in common with large spacecraft telemetry systems. Extrapolating from these, it can be readily appreciated that adaptive processing systems having the ability to meet changing situations will make the necessary diagnoses and evaluations available faster and more economically.

Beyond this, supersonic transports must have complete and reliable atmospheric weather information before takeoff in order to obtain optimum cruising altitudes and routes to planned destinations, so that they can avoid the heavy economic penalties involved in carrying excess fuel.

ADAPTIVE TELEMETRY CHARACTERIZATION

In a broad sense, any system capable of changing or adjusting to meet the requirements of a different condition can be called adaptive. Accordingly, any self-adjusting mechanism (such as AGC or AFC systems) would be included as well as other conventional open or closed loop feedback systems. Indeed, the entire field of biological systems, including man, would be included with the overriding criterion of self-preservation. Such a broad classification is not very useful for the purposes at hand. However, if we add the requirement that the systems of interest must modify themselves as the result of learning we are left with a much more restricted and interesting class of systems.

In particular, the specification and design of optimum (from the standpoint of a posteriori probability) Telemetry receivers are highly developed areas of engineering. That is, for nearly any given set of conditions, an optimum receiver can be hypothesized, analyzed, and at least nearly realized in practical form. In addition, adaptable receivers — that is, receivers that can adjust to optimum configurations for different sets of a priori conditions — are entirely practical. Such receivers imply an ability to make estimates on certain system parameters and to make adjustments in the decision structure which utilizes these estimates. If it is assumed that the system parameters are changing in an a priori unknown fashion, then such a receiver must be able to learn the condition of these system parameters; this requires the measurement or monitoring of performance against given criteria.

There are then three operations which can be required to delineate a more restricted as well as a more important class of adaptive telemetry systems. They are as follows:

(1) Monitoring the system performance by measurements against given performance criteria and relating these measurements to certain system parameters. Here the monitoring procedure refers to deriving certain parameters from either the data source at the transmitter or the observed data at the receiver for application to the learning process.

(2) Learning or estimating the states of certain channel parameters and evaluating the quality of the received information. Here the estimation

is derived from the monitoring signals and a learning process may be employed to change prior knowledge at the receiver.

(3) Modifying or adjusting parameters in a way which optimizes the receiver structure as instructed through the learning process.

It should be observed that these requirements do not restrict the location of the functions. Thus, adaptive procedures may be applied at the transmitter location, in which case encoding equipment may be adapted to accommodate the changes in the data source. Alternately, the adaptive procedures may be applied at the receiver location, in which case decision functions may be adapted to accommodate changes in the transmitter operation or changes in the transmission medium. A combination of adaptive procedures can be applied at both the transmitter and receiver locations with or without feedback. In the latter case, the ultimate user of the telemetry data might well be included in the system, perhaps to perform part of the learning and modifying functions.

Session I

ADAPTIVE PREDICTION AND DATA COMPRESSION

Chairman A. Balakrishnan
UCLA

Contributors L. Davisson
Princeton

R. Kutz
GSFC

T. Anderson
JPL

I. Eisenberger
JPL

W. Lushbaugh
JPL

E. Posner
JPL

1. ADAPTIVE METHODS IN COMMUNICATION SYSTEMS

A. V. Balakrishnan

University of California
Los Angeles, California

There has been much activity in the past concerning adaptive control systems; however, adaptive methods in communication theory are of more recent origin (References 1 to 4). The introduction of adaptive methods is a measure of sophistication in systems design made possible by the availability of high-speed large-scale digital computers. In communication systems (as well as in radar and data-processing systems generally), Earth-based receiver complexes already include fairly elaborate computing facilities, often as an integral part of the receiver, as, for example, in data transmission systems or space communication systems. The addition of some extra logic or arithmetic operations to accommodate adaptive methods is no longer a source of major difficulty. In fact, processing capabilities are already outstripping extant theory in many areas outside of communications, such as in geophysical prospecting, that can take advantage of them. In short, the need for adaptive methods cannot be overemphasized for optimal operations that are free from preset constraints and can adapt to changing data or system conditions.

The present paper treats the optimization theory for adaptive systems that involve the transmission and/or reception of information. After examining the features that define adaptivity, three specific applications are discussed: The adaptive feature in the transmitter, that in the receiver, and that in both receiver and transmitter. In the first case, a PCM transmitter is designed to minimize average transmitter power directly by adapting it to the probable state. The second application is a diversity receiver which seeks to minimize error by adapting to various signal strengths. The third is an adaptive data compression system that exploits data redundancy by utilizing an adaptive data predictor; the band-width reduction is initiated only if the data warrant it.

These examples also serve to illustrate the problems that still remain before a general theory of adaptive optimization can evolve. One of the main difficulties is the quantitative evaluation of adaptive systems.

DEFINING FEATURES OF ADAPTIVE SYSTEMS

An adaptive system is distinguished primarily by a learning mode that is either explicit or implicit. However, to be truly adaptive it must have other

features as well. The following interpretations of dictionary definitions serve to illustrate this point: Adaptive implies modification to meet new conditions; and adjusting implies a bringing into as exact or close a correspondence as exists between the parts of a mechanism, but suggests more tact or more ingenuity in the agent. Actually, it is impossible to draw a sharp distinction between what is adaptive and what is not. The word adaptive does signify something more than self-adjusting mechanisms (such as an AGC or AFC system where there are no conditions not envisaged at the time of design). Without getting involved in semantics, one can say that an adaptive system is one which monitors its own performance; when new conditions arise that degrade the performance, the system learns what they are and changes its structure accordingly.

In a communication system the new conditions might stem from a priori unknowable channel parameter changes. These parameter changes have to be learned or measured and appropriate changes must be made in the system to maintain a prescribed level of fidelity, or other chosen criterion. Let us consider some simple examples. The usual Shannon-Fano noiseless coding for minimal length needs a complete specification of the message probabilities. These probabilities are unknown and have to be measured. Moreover, if the probabilities are also subject to change, there must be a learning phase, whether it is separate from the coding phase or not. One cannot call such a system adaptive since no monitoring of the performance is involved. In a two-way or feedback communication system (such as a data transmission system) the receiver monitors the quality of performance (with an error-detecting code) and the transmitter adjusts by reducing the transmission rate or stopping the transmission altogether. Since there is no learning involved, this is not quite adaptive by the definition given above. The learning phase is triggered by a degradation in performance. A radar system that simply measures the clutter or reverberation parameters and then changes the detection circuitry parameters is not quite adaptive either, since there is no self-monitoring criterion. If the transmission system learns or measures channel characteristics and changes the mode of transmission accordingly to restore the performance quality, it would be a truly adaptive system.

Actually, the term "adaptive" is beginning to be used whenever there is a learning or measurement feature (implicit or explicit) even if there is no self-monitoring. (The terms "self-regulating" or "self-adjusting" are used for systems with the monitoring feature but no learning feature.)

This paper is concerned with only learning or adaptive systems. In keeping with this, adaptive communication systems are categorized as: (1) Transmitter adaptive (adaptive methods at the transmitter), (2) receiver adaptive (adaptive methods at the receiver), and (3) both receiver and transmitter adaptive. The

classification is based upon whether the primary adaptive mechanism is at the transmitter or at the receiver, even though a measure of complementary adaptivity may be involved at both ends. The discussions that follow are intended to illustrate the formulation of the problems that arise.

TRANSMITTER ADAPTIVE SYSTEMS

An adaptive system in which the major adaptive function is at the transmitter (but without the performance monitoring feature) will be considered first. A binary source is considered in which successive bits are statistically independent and the need is to transmit the information coherently using, as usual, $A_1 \cos \omega_c t$ or $A_2 \cos (\omega_c t + \theta)$, $0 \leqq t \leqq T$. If we use a coherent receiver and the channel noise is characterized by additive white Gaussian noise, it is readily seen (Reference 5) that $\theta = \pi$ should be chosen without regard to A_1 and A_2. It is also assumed that state one has probability P_1 and that this is known to the transmitter and receiver. Where $\omega_c T >> 1$, the average transmitted power is

$$P_1 \frac{A_1^2}{2} + P_2 \frac{A_2^2}{2} \quad . \tag{1}$$

The probability of error when the optimal receiver is used is an involved problem because, ordinarily, $P_1 = 1/2$ in the optimum receiver design. Among the reasons for this, the all-important one is that the receiver cannot do better than to minimax; hence the receiver threshold is $(A_1 - A_2)/2$ and the probability of error is decreased by increasing $A_1 + A_2$. It is then assumed that $A_1 + A_2$ is fixed to provide a desired error rate; e.g., let $A_1 + A_2 = A$. The transmitter is designed to provide the minimum output power consistent with the minimum acceptable error. The average power given by Equation 1 can be minimized for a fixed value of A by taking $A_1 = (1 - P_1)A$ and $A_2 = P_1 A$, yielding the minimum $1/2\ P_1(1 - P_1)A$. Such a selection is obviously in accord with one's intuition. Since more power is provided for the less likely state, the power reduction can be substantial. A closer examination of the notion of average power is now in order. A phase-average has been taken and, therefore, the law of large numbers must be invoked to interpret it as a time average

$$\frac{1}{nT} \int_0^{nT} A(t)^2 \, dt = \frac{1}{n} \sum_1^n \frac{A_i^2}{2} \doteq \frac{1}{2} E(A_1^2).$$

By the approach described the average power is reduced while the same error rate is maintained. The only difficulty is that the probability P_1 in any

practical communication system is unknown and must be evaluated en route. Here the system has to adapt to the particular signal structure which cannot be predetermined. If the view is taken that the probability P_1 has to be evaluated first and the amplitudes A_i chosen accordingly, then the problem of determining what constitutes an optimum estimate or, more properly, what constitutes an optimum procedure arises. It is clear that it is better to seek the overall optimum directly. If it is assumed that the decisions are to be based on an N-chain of digits S_i, where $S_i = 1$ or 2, and (A_i) denotes the chosen amplitudes (based on the past N digits so that $A_n = f(S, \ldots, S_{n-N})$, where $A = f(1, S_{n-1}, \ldots, S_{n-N}) + f(2, S_{n-1}, \ldots, S_{n-N})$), then to minimize the average power, when the message probability structure is known,

$$E\left(A_n^2\right) = \sum \left[f\left(1, S_{n-1}, \ldots, S_{n-N}\right)^2 P_1 + f\left(2, S_{n-1}, \ldots, S_{n-N}\right)^2 \left(1 - P_1\right) \right] P\left(S_{n-1}, \ldots, S_{n-N}\right).$$

Here $P_1 = P(1/S_{n-1}, \ldots, 1/S_{n-N})$ yields the optimum choice,

$$f\left(1, S_{n-1}, \ldots, S_{n-N}\right) = A\left[1 - P\left(1/S_{n-1}, \ldots, 1/S_{n-N}\right)\right].$$

When the bits are independent, this agrees with the heuristic choice discussed above. The main difficulty is that these probabilities are unknown and must be estimated. To proceed in this case, it is assumed that there is an unknown parameter (multidimensional, perhaps) such that

$$\int P\left(S_n, \ldots, S_{n-N} / \theta\right) \rho(\theta)\, d\theta = \rho\left(S_n, \ldots, S_{n-N}\right),$$

where $\rho(S_n, \ldots S_{n-N}\theta)$ is known for every value of θ, but the distribution $\rho(\theta)$ is unknown although negotiable (in a manner to be discussed). If it is assumed that θ is a random variable, and the N-chain $[S_{n-1}, S_{n-2}, \ldots, S_{n-N}]$ is denoted by the vector S, then the dependence on θ can be shown by

$$E\left(A_n^2\right) = \int \sum \left[f(1, S)^2\, P(1, S/\theta) + f(2, S)^2\, P(2, S/\theta) \right] \rho(\theta)\, d\theta$$

$$= \int \sum \left[f(1, S)^2 \, P(1/S, \theta) + f(2, S)^2 \, P(2/S, \theta) \right] P(S/\theta) \, d\theta,$$

so that the optimal choice can be expressed

$$f(1, S) = A \frac{\int P(2, S/\theta) \rho(\theta) \, d\theta}{\int P(S/\theta) \rho(\theta) \, d\theta}$$

and

$$f(2, S) = A \frac{\int P(1, S/\theta) \rho(\theta) \, d\theta}{\int P(S/\theta) \rho(\theta) \, d\theta} .$$

The similarity of the present problem to decision-theoretic problems allows the methods and procedures of decision theory to be utilized in making an evaluation of $\rho(\theta)$. First, it is noted that the optimal choice can be written

$$f(1, S) = A \int P(2/S, \theta) P(\theta/S) \, d\theta$$

and

$$f(2, S) = A \int P(1/S, \theta) P(\theta/S) \, d\theta.$$

These are conditional expectations of the probabilities $P(2/S, \theta)$ and $P(1/S, \theta)$; hence, they are also the best estimates of these quantities (in the mean square sense). In other words, when the probabilities $P(1/S)$ and $P(2/S)$ are not known, the theory recommends the use of the best mean square estimates, however the choice of $\rho(\theta)$ may still be made. Maximum-likelihood estimates can be used which, in a sense, eliminate the a priori density problem. Thus, by utilizing the value of θ that maximizes $P(S/\theta)$, say $\hat{\theta}$,

$$f(1, S) = P \; 2/S, \hat{\theta}).$$

For large values of N (from the asymptotic theory) it is expected that these various estimates will not differ much. Thus, the example is considered where the successive bits are independent for each value of θ, so that

$$f(2, S) = \frac{\int \theta P(S/\theta)\, \rho(\theta)\, d\theta}{\int P(S/\theta)\, \rho(\theta)\, d\theta}$$

If $0 \leqq \rho(\theta) \leqq 1$, $0 \leqq \theta \leqq 1$, and $P(S/\theta) = \theta^{\nu}(1 - \theta)^{N-\nu}$ corresponding to ν 'one' states is chosen, then

$$f(2, S) = \frac{\int_0^1 \theta^{\nu+1}(1 - \theta)^{N-\nu}\, d\theta}{\int_0^1 \theta^{\nu}(1 - \theta)^{N-\nu}\, d\theta} = \frac{\nu + 1}{N + 2} .$$

On the other hand, if the maximum likelihood estimate is used, $f(2, S) = \nu/N$, which substantiates our observation.

The minimax estimates can also be obtained as in decision theory, but due allowance must be made for the fact that this may be overly pessimistic. Again, asymptotically, many of these estimates may be equivalent, but it is not clear what the practical implication of asymptotic optimality is.

It should be noted that successive digits were assumed to be independent. This is of course an ad hoc assumption as is the parametric model. Clearly, little more than a statement of the problem has been achieved. Since the object here was to point out some of the general features and some of the theory that needs to be developed, the groundwork has been laid for the next example.

RECEIVER ADAPTIVE SYSTEMS

This second example considers a system in which the adaptive learning feature is at the receiver, but, as with the previous example, the nature of the theory involved will be treated without going into specific details. For this example a PCM system employing "space diversity" (i.e., several receivers with n of them tuned to the same transmitter) will suffice. In this system the received signals at the i^{th} station can be expressed by

$$f_i(t) = k_i\, s(t) + N_i(t), \qquad 0 \leq t \leq T, \qquad i = 1, \ldots, n ,$$

where N_i (t) is the noise (for instance additive Gaussian) at each receiver, s(t) is the information-carrying waveform

$$s(t) = s_k(t), \; k = 1, \ldots, M$$

which is known at each receiver, and k_i is the amplitude factor which may vary from receiver to receiver. Where the (k_i), or the ratios of the various k_i's are known, the optimal receiver characteristics are derived that minimize the probability of error and use all of the received signal $\left(f_i(t)\right)$. The point is that it is possible to do better by a suitable combination of the $f_i(t)$ than by individual processing (detecting) at each receiver and then using some kind of majority logic to decide on which of the signals $\left(s_k(t)\right)$ was transmitted. The optimal receiver structure can be easily derived by considering $f_i(t)$ as a "multiple" process and setting

$$F(t) = \text{col}\left[f_1(t), \ldots, f_n(t)\right],$$

$$S(t) = \text{col}\left[k_1 s(t), k_2 s(t), \ldots, k_n s(t)\right],$$

$$S_i(t) = \text{col}\left[k_1 s_2(t), k_2 s_i(t), \ldots, k_n s_i(t)\right],$$

$$N(t) = \text{col}\left[N_1(t), \ldots, N_n(t)\right].$$

If we let R(s, t) be the covariance matrix function, then

$$R(s, t) = E\left[N(s)N(t)^*\right]$$

where the ( )* denotes a conjugate transpose. If it is assumed that the signals $s_k(t)$ are equally likely, it can be shown (Reference 6) that the k^{th} signal should be chosen if

$$\min_i \int_0^T h_i(t)^* S_i(t)\,dt - 2\int_0^T F(t)^* h_i(t)\,dt$$

$$= \int_0^T h_k(t)^* S_k(t)\,dt - 2\int_0^T F(t)^* h_k(t)\,dt,$$

where

$$\int_0^T R(s, t) h_i(t)\, dt = S_i(s).$$

The simplest case occurs when

$$R(s, t) = \Phi I \delta(s - t),$$

where I is the unit matrix and corresponds to the case where the noises are pure white and independent with the same spectral density. A somewhat more general case is

$$R(s, t) = D\delta(s - t),$$

where D is a diagonal matrix with positive entries which, perhaps, correspond to different noise temperatures at the receivers. In this case

$$h_i(t) = \text{col}\left(\frac{k_1}{d_1} s_i(t), \frac{k_2}{d_2} s_i(t), \ldots, \frac{k_n}{d_n} s_i(t)\right),$$

where (d_i) are the diagonal elements of D. Also, if it is assumed that all signals have the same power and are consistent with the assumption that all signals are equally likely, then

$$\int_0^T h_i(t) * S_i(t)\, dt = \left(\sum_1^n \frac{k_i}{d_i}\right) PT,$$

where

$$\frac{1}{T} \int_0^T s_i(t)^2\, dt = P.$$

By substituting these results into the decision rules, the expressions reduce to

$$\max_i \int_0^T f(t) * D^{-1} S_j(t)\, dt \;=\; \int_0^T f(t) * D^{-1} S_k(t)\, dt$$

or

$$\max_i \int_0^T s_j(t) \left(\sum_1^n \frac{k_i f_i(t)}{d_i} \right) dt \;=\; \int_0^T s_k(t) \left(\sum_1^n \frac{k_i f_k(t)}{d_i} \right) dt \,.$$

In other words, each receiver waveform $f_i(t)$ has to be weighted by the corresponding signal-to-noise ratio k_i/d_i.

For the adaptive feature, it is noted that the (k_i) or their ratios in a diversity system, are usually unknown, or are subject to variation. Since the noise temperatures are usually known, the (d_i) are known. Thus, a learning phase is included where the (k_i) are estimated and then an "adaptive" phase is included where the weighting factors are adjusted accordingly. Such a problem for the case of binary signals and white noise has been treated by Price (Reference 1). One such method of estimation is the case where it is possible, by a suitable operation at the receiver, to change $S_i(t)$ to one parameter, regardless of what i may be. For example, if the system is polyphase (one fixed phase angle $2\pi/k$, where $k = 1, \ldots, M$ corresponds to each signal), then, by multiplying the phase by the factor k, the modulation uncertainty is removed.

The problem of estimating the parameters (k_1) by "maximum likelihood" or similar criterion remains. In general, if it is assumed that all the waveforms $s_i(t)$ are equally likely and that the (k_i) are mutually independent, then by maximizing the likelihood function the setting for each i, becomes

$$\frac{\partial}{\partial k} \log \left(\sum_{j=1}^{M} \frac{1}{M} \exp \left\{ - \frac{1}{2d_i} \int_0^T \left[f_i(t) - k\, s_j(t) \right]^2 dt \right\} \right) \;=\; 0,$$

or by taking

$$\sum_{j=1}^{M} \int_0^T \left[f_i(t) - k s_j(t)\right] s_j(t)\, dt \exp\left\{-\frac{1}{2d_i} \int_0^T \left[f_i(t) - k s_j(t)\right]^2 dt\right\} = 0$$

with

$$\int_0^T s_j(t)^2\, dt = PT,$$

the setting becomes

$$\sum_{j=1}^{M} \left[\int_0^T f_i(t)\, s_j(t)\, dt\right] \exp\left[+\frac{k}{d_i} \int_0^T f_i(t)\, s_j(t)\, dt\right]$$

$$= k \sum_{j=1}^{M} PT \exp\left[\frac{k}{d_i} \int_0^T f_i(t) s_j(t) dt\right].$$

Actually, the adaptive problem has just begun. An optimum estimate of (k_i) should be based on a minimum probability of error by using any a priori statistics available. Any optimum procedure devised would have to be evaluated in terms of the probability of error that results. Extension to noncoherent transmission is a natural one since practical systems utilizing adaptive estimation (such as optical communication systems where diversity can be of use to offset the effect of multiple scattering) would be noncoherent.

ADAPTIVE DATA COMPRESSION SYSTEM

As a final example, a system is considered that is adaptive at both the transmitter and the receiver. This system is more truly adaptive in the sense that it features a self-monitoring function in addition to the learning and adjusting functions. In contrast to the preceding section, nonparametric methods will now be used.

Let $x(t)$ represent a continuous or discrete parameter source. It is customary to assume that $x(t)$ can then be regarded (at least for analytical purposes)

as a stochastic process. If the time-parameter t is continuous, then it is possible to assume that the expression

$$x(t) = \int_{-B}^{B} e^{2\pi i f t} \, d\psi(f)$$

is adequate (depending on whether we adopt the stochastic or nonstochastic viewpoint) for a sufficiently large bandwidth B. By the well-known "sampling principle" one can represent the continuous waveform by using periodic samples taken at $t = \frac{n}{2B}$. In most communication systems using sampled data of this kind, the sampling rate is determined by the nominal, highest, or cutoff frequency B expected in the data. However, in many kinds of data (such as in space telemetry data) the actual cutoff frequency will be small most of the time, so there is no need to sample at the nominal rate of 2B; that is, it is possible to "compress" the data sampling rate or channel bandwidth. A method of achieving such compression in PCM systems is to exploit the redundancy or predictability of the data. Thus, a "predictor" is employed at the transmitter which predicts the data at time $t + \Delta$ based on the data up to time t. The actual value observed is then compared with that predicted. If the difference exceeds (in absolute value) a preset threshold, then the actual sample is transmitted. If it is below the threshold, a prearranged code word using one digit could be transmitted instead of the m digits; this reduces the number of digits transmitted per second. A comma-free code may be used to sort the two kinds of words unambiguously. The adaptive or learning feature is part of the predictor mechanism. The predictor is not operative until the threshold is exceeded and thus exhibits the self-monitoring feature. The predictor itself is based on a learning phase and the prediction operator is adjusted accordingly. For details on the predictor itself, see Reference 2. The significance of the adaptive prediction feature lies in the fact that no a priori statistics or other assumptions concerning the data are required and, in particular, that there may be periods in the data where prediction (to the quality set) may not be possible.

The basic prediction philosophy may be indicated briefly. A waveform of duration T (a function $x(t)$, $0 \leq t \leq T$) is given and the prediction of the value at time $T + \Delta$ is required. Any prediction operation is to be based on this data alone since no additional a priori knowledge is available. This is, of course, an ancient problem; however, we wish to treat it strictly in the telemetry data processing context and note two major points of view in dealing with it. One which may be considered the "numerical analysis" point of view consists in assuming that any physically realized waveform must be analytic and can be approximated by polynomials. The data may thus be "fitted" to a polynomial of a

high degree and then used to "predict" the future values. The other and more recent view is the statistical view in which the data are assumed to be a finite sample of a stationary stochastic process whose average properties, such as moments and/or distributions, are known or calculable from the data. For a process of a given description, the well developed mean square prediction theory can be applied. Perhaps the main advantage with this view is that it gives a quantitative, although theoretical, notion of the error in prediction. The problem of measuring the average statistics from a finite sample can be quite delicate, however. In the polynomial fitting method, the interpretation of the prediction error (which is the crucial point) is linked with what degree polynomial to be used and what portion of the data is to be fitted and is therefore more nebulous. Moreover, as ordinarily used, the fitting operations on the data are linear.

A rationale is adopted for prediction which is free from a priori assumptions concerning the "model." In a general sense, what is involved in both of the above methods is (1) "modelmaking", consistent with the data and (2) using the numbers derived from the model to perform some optimal operations. If an understanding of the mechanism generating the model is desired, the first step is essential. If prediction is wanted, then it shall be shown that it is possible to proceed directly (and hence more optimally) to the best prediction without the intermediate step of modelmaking. It may be, of course, that several philosophies lead to the same operations on the data. Even here, the present method offers some practical advantages. It is only natural to use the philosophy that requires the least number of a priori assumptions.

Any prediction is an operation or operator on part of the data, and in this case the finite part is all that is available. The main point of departure in the view being taken is that if a prediction operator is obtained which, based on all the available input data, functions optimally in the immediate past relative to the point where the prediction is required, the most meaningful solution to the prediction problem is achieved. Thus, the only basis on which an a priori judgement can be made on any prediction method is to "back off" slightly from the present and compare the actual available data with the predicted value by use of the given prediction operator. This can be formulated analytically as follows: Let the total available data be described as a function $x(t)$, $0 \leqq t \leqq T$, and let it be required to predict the value at $T + \Delta$, where Δ is small fraction of t that is greater than zero. Next consider the data in the interval $0 < t < T - \Delta$. If for any t_0 in this interval a "prediction" of the function value Δ is made from the past values up to t_0, and the predicted value is given by

$$x^*(t_0 + \Delta)$$

and the error can be explicitly observed as

$$x^*(t_0 + \Delta) - x(t_0 + \Delta).$$

At this point it should be noted that the general prediction operator will be a function of a finite segment of the past. If the length of this segment is denoted by S, then

$$x^*(t_0 + \Delta) = O\left[x(\sigma) \text{ where } (t_0 - S) < \sigma\ t_0\right], \tag{2}$$

where O represents the prediction operator. The error criterion must be specified next. Here the mean square error has been chosen, because it is simpler analytically, is almost universally used, and makes possible a comparison with other methods. The kind of solution presented is based on successive approximation as opposed to an analytic solution or to other measures of error which risk greater complexity. As far as the rationale of the method is concerned, this is more a matter of detail than principle. Optimality is determined by using

$$\frac{1}{T - \Delta - L} \int_L^{T-\Delta} \left[x^*(t + \Delta) - x(t + \Delta)\right]^2 dt = \epsilon^2, \tag{3}$$

where $L \geq S$, and by continuing the procedure on the basis that the operator O which minimizes Equation 3 is the best. Equation 3 can be generalized to

$$\frac{1}{T - \Delta - L} \int_L^{T-\Delta} \rho(t) C\left[x^*(t + \Delta) - x(t + \Delta)\right] dt, \tag{4}$$

where $\rho(t)$ is a positive weight function and C ( ) is a symmetric positive cost function. Before the problem of determining the optimal operator O is considered, an important consistency principle should be noted. If the data are regarded as one long sample of an ergodic process, then, in the statistical sense, Equation 3 yields the optimal operator exactly, however, it is not necessary to make any such assumption concerning the data, nor to compute the average statistics first. The point is, that although the data may not be enough for determining the spectrum, they may be quite adequate for the prediction itself. Unlike the polynomial fitting, the operations on the data can be as nonlinear as necessary, and at the same time Equation 3 can be normalized to

$$\epsilon^2 \Big/ \left(\frac{1}{T - \Delta - L} \int_L^{T-\Delta} x(t + \Delta)^2 \, dt \right), \tag{5}$$

thus yielding a quantitative measure of the prediction error by which to judge prediction how good the prediction will be. For actual methods of determining the optimum predictor see Reference 2.

In the system, the receiver is assumed to perform a prediction operation similar to that performed by the transmitter; that is, the receiver predicts the values of the nontransmitted data using the transmitted data as well as predicted sections of the data. This would mean that the receiver sets the same level of possible prediction operator complexity as the transmitter does. Specifically, the transmitter itself has to base its prediction upon using, as necessary, the predicted data points that did not exceed the threshold error.

So far, the optimal adaptive structure has been determined, but the system must still be evaluated. It is natural to ask how much compression can be obtained in this way on the average. Here some structure must be postulated for the data source and then the resulting compression must be calculated using the adaptive system. The following briefly indicates what an analysis of this type involves. To simplify the analysis, consideration of the prediction operation in the transmitter is based on the actual samples only. The data samples are denoted by (x_n). It is assumed that the data can be taken as a stationary stochastic process. It is further assumed that the process is Gaussian since the maximum prediction error (and hence the minimum compression) occurs in this case and the prediction operation includes only the linear operations. When the case where the adaptive predictor is also constrained to be linear is considered, the optimal filterweights α_j are determined by minimizing

$$\frac{1}{N} \sum_{-N+1}^{0} \left(x_n - \sum_{1}^{m} \alpha_j x_{n-j} \right)^2, \tag{6}$$

where the available past data consist of N samples and the prediction is based on "m" samples (considering the "analog" method above). The corresponding (squared) error is then

$$\epsilon_0^2\ [mN] \;=\; \left(x_1 - \sum_1^m \alpha_j\, x_{1-j}\right)^2.$$

The transmission is based on Equation 7 exceeding a threshold "t." Hence, the statistics of Equation 7 must first be determined, noting that the optimal (α_k) that minimize Equation 6 will satisfy

$$\sum_{-N+1}^{0} x_n\, x_{n-k} \;=\; \sum_j \;\; \sum_{n=-N+1}^{0} \alpha_j\, x_{n-j}\, x_{n-k},\quad k \;=\; 1,\ \ldots,\ m.$$

If Y_n is the N-column vector $(x_{+n},\ x_{+n-1},\ \cdots\ x_{n-N+1})$, then the minimum of Equation 6 becomes

$$\frac{1}{N}\ \frac{D_{m+1}}{D_m}\ ,$$

where D_{m+1} is the determinant of the $m + 1$ by $m + 1$ matrix with entries

$$Y_i \cdot Y_j,\ i, j \;=\; 0,\ -1,\ \ldots,\ -m$$

and D_m is the determinant of the m by m matrix with entries

$$Y_i \cdot Y_j,\ i, j \;=\; -1,\ -2,\ \ldots,\ -m.$$

On the other hand, the error is of interest (Equation 7) which is

$$\epsilon_0^2\ [m;N] \;=\; \left(\frac{D'_{m+1}}{D_m}\right)^2, \tag{8}$$

where D'_{m+1} has the first row

$$x_1,\ x_0,\ x_{-1},\ \ldots,\ x_{-m+1}$$

and is otherwise the same as D_{m+1}. The statistics of Equation 8 are of primary interest. In this case it is necessary to know

$$\Pr.\left[\epsilon_0^2\ [m;N] \geq t\right],$$

which is the probability of threshold excedence and relates directly to the attainable compression. Actually Equation 8 determines only the "open-loop" error. To evaluate the complete system, the "closed-loop" operation must be considered, where actual samples are replaced by predicted samples when the former fall within the set threshold. Needless to say, the analysis tends to get involved even for the Gaussian model.

It is equally important to know what the "quality" of the compressed data will be. The cross-correlation coefficient or other equivalent quantitative index may be used and evaluated as a function of t, m, and N. A complete system evaluation would also require that the channel noise be taken into account. Such an analytic evaluation is necessary to provide a logical basis for choosing optimum design parameters. Indeed, if the channel noise characteristics are known, then appropriate coding or signal selection has to be considered in arriving at the total compression possible.

In all of these cases cited, the adaptive theory winds up with an "optimum" procedure. It is necessary to be able to compare such procedures on a quantitative basis. To evaluate how good the system actually is, however, it is necessary to specify the class of input or system parameters, or their statistics if they are regarded as randomly varying. This has many difficulties since, by its very adaptive nature, the system has to cope with conditions which cannot be preassigned. One way out, as indicated, is that of asymptotic optimality; however, this is not entirely satisfactory either.

In conclusion, it should be noted that adaptive methods in communication theory are still in their formative stages. Before any general theory can be evolved some specific problems, such as the ones mentioned, need to be explored in detail. The scope of the theory is quite far-reaching and the examples discussed are intended primarily to state some specifics rather than vague generalities.

REFERENCES

1. Price, R., "Error Probabilities for Multichannel Reception of Binary Signals," TR No. 258, Lincoln Lab., 1962.

2. Balakrishnan, A. V., "An Adaptive Nonlinear Data Predictor," Proc. National Telemetering Conference, May 23-25, 1962, Washington, D.C.), New York, N. Y.: The Inst. of Radio Engineers, Paper 6-5, Vol. II, pp. 1-15.

3. Weaver, C. S., "Adaptive Communication Filtering," Int. Radio Eng. Trans. on Information Theory, IT-8(5):S169-S178, September 1962.

4. Balakrishnan, A. V., "Prediction and Filtering" in "Report on Progress in Information Theory in the U.S.A., 1960-1963" IEEE Trans. on Information Theory, IT-9(4):237-240, October 1963.

5. Balakrishnan, A. V., "A General Theory of Non-Linear Estimation Problems in Control Systems," J. Math. Analysis and Applications, 8(1):4-30, February 1964.

6. Balakrishnan, A. V., "Estimation and Detection Theory for Multiple Stochastic processes," J. Math. Analysis and Applications, 1(3):386-410, December 1960.

2. ADAPTIVE PREDICTION OF TIME SERIES

L. D. Davisson

Princeton University
Princeton, New Jersey

INTRODUCTION

The optimum operations for the "pure" prediction (no noise) of a discrete time series $[s_i]$ with respect to a mean square error cost function are well known when the data probability distributions are given. Unfortunately, there are many practical situations (References 1, 2, and 3) where it is not possible to specify these probability distributions and an adaptive technique is required in order to approach the minimum variance. There is no universally agreed upon adaptive procedure, but most systems rely upon the minimization of an appropriate quadratic form in the data close to the point to be predicted. In one such system (References 1 through 4), the predictor form is taken to be a linear weighting of the values immediately preceding the point to be predicted, the number of values (memory) being fixed a priori. The sample mean square error (fitting error) is minimized with respect to the weights in finding the adaptive predictor. In Reference 4 it is shown that the mean square prediction error for stationary Gaussian data with square integrable spectral density and mean zero is

$$\sigma^2\,[M, N] = \sigma^2\,[M, \infty]\left(1 + \frac{M}{N}\right) + o\left(\frac{M}{N}\right),$$

where M is the memory and N (learning period) is the number of predictions used in calculating the sample mean square error. It is desirable to keep N small in order to minimize the effect of nonstationarities. On the other hand, it can be seen that the variance does not generally decrease monotonically with M for fixed values of $N < \infty$ since $\sigma^2\,[M, \infty]$ is a decreasing function of M which approaches a constant and $\left(1 + \frac{M}{N}\right)$ is an increasing function of M. Thus, there is an $M < \infty$ such that $\sigma^2\,[M,N]$ is minimized (assuming $\sigma^2\,[\infty,\infty] > 0$). If something is known about the behavior of $\sigma^2\,[M,\infty]$, the optimum value of M can be estimated a priori. Usually this is not the case.

This paper is concerned with the estimation of $\sigma^2[M,N]$ from the data for the same Gaussian data model. Unfortunately, the fitting error is a poor estimate since it is monotonic and decreases with M even though $\sigma^2[M,N]$ is not monotonic. However, an appropriately weighted function of the fitting error can be used. On the basis of this estimate it is then possible to specify the best value of M for fixed values of N and then to specify a value of N so that the factor $\sigma^2[M,N]/\sigma^2[\infty,\infty]$ is sufficiently small. The values of M/N considered are "intermediate" in the sense that terms of $o(M/N)$ in asymptotic expansions can be neglected whereas the M/N term cannot. Specifically, it is shown that if $F^2(M,N)$ is the fitting error or sample prediction error variance then

$$E\left[F^2(M,N)\right] = \sigma^2[M,\infty]\left(1-\frac{M}{N}\right)+\frac{o(M)}{N}e+\frac{o(M)}{N} \quad .$$

Thus an asymptotically unbiased estimate of $\sigma^2[M,N]$ is

$$\hat{\sigma}^2[M,N] = F^2(M,N)\frac{1+\frac{M}{N}}{1-\frac{M}{N}} \quad .$$

The minimization of this quantity as a function of M over all values of M up to some maximum value M_{max} such that $M_{max}/N << 1$ provides an estimate of the optimum memory M_o. The quantity

$$\hat{\sigma}^2\left[M_{max},\infty\right] \quad F^2(M_{max},N) \cdot \frac{1}{1-\frac{M_{max}}{N}}$$

gives an estimate of $\sigma^2\left[M_{max},\infty\right]$. Now the learning period is adjusted to a value N_o so that

$$\frac{\hat{\sigma}^2\left[M_o,N_o\right]}{\hat{\sigma}^2\left[M_{max},\infty\right]} \leq 1+\epsilon \ ,$$

where ϵ is an allowed deviation of the ratio of the mean square error to the minimum variance from unity. This procedure is motivated by the desire to bring the mean square error within ϵ of the optimum while keeping N_o as small as possible. No optimality is claimed for the procedure except possibly in the

limit as $\epsilon \to o$. However, to the author's knowledge there is no other available procedure to accomplish this. The nonlinear relations between the data and the predictor when the data probability distributions are unknown make an exact analysis extremely difficult if not impossible.

The results of computer simulations are presented in a later section to verify the effectiveness of the procedure.

DESCRIPTION OF PREDICTOR

As noted in the introduction, linear predictions of the time series $[s_i]$ are sought in the form

$$\hat{s}_i = S_i' B_{MN},$$

where B_{MN} and S_i are M column vectors

$$S_i = \{s_{i-j},\ j = 1, \cdots, M\},$$

and B_{MN} is a weight vector determined adaptively for each value of the predictor memory M from sample predictions of the N points preceding the point to be predicted (taken for convenience to be s_0, the value at the time origin). The sample error variance is then minimized to determine B_{MN}. The resulting minimum is the fitting error

$$F^2(M,N) = \min_{B_{MN}} \frac{1}{N} \sum_{i=1}^{N} (s_{-i} - S_{-i}' B_{MN})^2 . \qquad (1)$$

This quadratic form has a minimum for

$$\frac{1}{N} \sum_{i=1}^{N} s_{-i} (s_{-i} - S_{-i}' B_{MN}) = 0 .$$

For convenience, this is expressed in the matrix form

$$g_{MN} = r_{MN} B_{MN},$$

where g_{MN} is a column vector and r_{MN} is a matrix given by

$$g_{MN} = \frac{1}{N} \sum_{i=1}^{N} S_{-i} s_{-i}$$

and (2)

$$r_{MN} = \frac{1}{N} \sum_{i=1}^{N} S_{-i} S'_{-i} \quad .$$

Thus (assuming r_{MN} is nonsingular),

$$B_{MN} = r_{MN}^{-1} g_{MN} \tag{3}$$

Now, given the fitting error $F^2(M,N)$, an adaptive procedure must be found by which M and N can be determined. This cannot be done from $F^2(M,N)$ directly, because, although it converges to $\sigma^2(M,\infty)$ as $N \to \infty$, for fixed values of N it decreases monotonically with M (as the number of degrees of freedom increase) even though the mean square prediction error does not (Reference 4). The determination of the adaptive procedure must be postponed until the mean fitting error $E[F^2(M,N)]$ is evaluated.

MEAN FITTING ERROR

The analysis of the mean fitting error proceeds as in Reference 4. The coefficient vector B_{MN} is a highly nonlinear function of the data as shown by Equation 3. However, asymptotically* (References 4 and 5), if the process spectral density is square integrable,

$$B_{MN} = r_{MN}^{-1} g_{MN} = R_M^{-1} G_M + R_M^{-1} \left(g_{MN} - r_{MN} R_M^{-1} G_M \right) + \epsilon_{MN} \ , \tag{4}$$

*To make this expression rigorously correct it is required to use a constraint in the form $||B_{MN}|| \leq CN^P$, $C > 0$, $P > 0$ to take care of singularities in r_{MN}. This is readily done by modifying the definition of Equation 3. Since Equation 1 is usually minimized by a gradient technique rather than by inverting r_{MN}, no practical difficulties are encountered in this definition.

where

$$R_M = E\left[r_{MN}\right] = \text{M by M covariance matrix of } s_i \ ,$$

$$G_M = E\left[g_{MN}\right] ,$$

and

$$E\left[\left|\left|\epsilon_{MN}\right|\right|^2\right] = o\left(\frac{M}{N}\right) .$$

From the square integrability of the process spectral density it follows that R_M is nonsingular for $M < \infty$ (Reference 4). Hence, $B_{M\infty} = R_M^{-1} G_M = \lim_{N\to\infty} r_{MN}^{-1} g_{MN}$ is the optimum set of coefficients. The second term in the expansion of Equation 4 represents the effects of "small" statistical fluctuations and ϵ_{MN} is the remaining error whose contribution goes to zero in the mean square sense faster than that of the second term as $M/N \to 0$. The mean square prediction is by definition

$$\sigma^2\left[M,N\right] = E\left[s_o - B_{MN}' S_o\right]^2 \ .$$

The establishment of its asymptotic value is outlined as a prelude to finding the mean fitting error. Details can be found in Reference 4.

This expectation can be expanded in terms of the minimum variance coefficients $B_{M\infty}$ as

$$\sigma^2\left[M,N\right] = E\left[s_o - S_o' B_{M\infty}\right]^2 + E\left\{2\left[s_o - S_o' B_{M\infty}\right]\left[S_o'\left(B_{M\infty} - B_{MN}\right)\right] + \left[S_o'\left(B_{M\infty} - B_{MN}\right)\right]^2\right\} .$$

As shown in Reference 4, by use of the asymptotic expansion of Equation 4, this reduces to

$$\sigma^2\left[M,N\right] = \sigma^2\left[M,\infty\right]\left(1 + \frac{M}{N}\right) + o\left(\frac{M}{N}\right) . \qquad (5)$$

The same method can be used to evaluate the expected fitting error. From Equation 1, by using the solution of Equation 3 for the coefficients the expected value of the fitting error is

$$E[F^2(M,N)] = E\left[\frac{1}{N}\sum_{i=1}^{N}\left(s_{-i}-S'_{-i}B_{MN}\right)^2\right]$$

$$= E\left[\frac{1}{N}\sum_{i=1}^{N} s_{-i}\left(s_{-i}-S'_{-i}B_{MN}\right)\right] \quad .$$

By expanding about the optimum coefficients and using Equation 4,

$$E[F^2(M,N)] = \frac{1}{N}\sum_{i=1}^{N}\left\{E\left[s_{-i}\left(s_{-i}-S'_{-i}B_{M\infty}\right)\right] - E\left[s_{-i}S'_{-i}\left(B_{MN}-B_{M\infty}\right)\right]\right\}$$

$$= \sigma^2[M,\infty] - E\left[g'_{MN}\left(B_{MN}-B_{M\infty}\right)\right]$$

$$= \sigma^2[M,\infty] - E\left[g'_{MN}r_{MN}^{-1}\left(g_{MN}-r_{MN}R_M^{-1}G_M\right)\right]$$

$$= \sigma^2[M,\infty] - E\left[\left(R_M^{-1}G_M + R_M^{-1}\left(g_{MN}-r_{MN}R_M^{-1}G_M\right) + \epsilon_{MN}\right)'\right.$$

$$\left.\left(g_{MN}-r_{MN}R_M^{-1}G_M\right)\right]$$

$$= \sigma^2[M,\infty] - E\left[\left(g_{MN}-r_{MN}R_M^{-1}G_M\right)' R_M^{-1}\left(g_{MN}-r_{MN}R_M^{-1}G_M\right)\right] + o\left(\frac{M}{N}\right) .$$

The expected value of the quadratic form in Equation 6 remains to be evaluated. Since $\{s_i\}$ is stationary, $E[s_i s_{i+j}] = E[s_o s_j]$. Therefore

$$E\left[\left(g_{MN}-r_{MN}R_M^{-1}G_M\right)' R_M^{-1}\left(g_{MN}-r_{MN}R_M^{-1}G_M\right)\right]$$

$$=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left[\left(s_{-i}-S_{-i}'B_{M\infty}\right) S_{-i}'R_M^{-1}S_{-k}\left(s_{-k}S_{-k}'B_{M\infty}\right)\right]$$

$$=\frac{1}{N}\sum_{i=-N}^{N}\frac{N-|i|}{N}E\left[\left(s_0-S_0'B_{M\infty}\right) S_0'R_M^{-1}S_i\left(s_i-S_i'B_{M\infty}\right)\right] .$$

From the square integrability of the process spectral density it follows that

$$\sum_{i=-N}^{N}\frac{N-|i|}{N}E\left[\left(s_0-S_0'B_{M\infty}\right) S_0'R_M^{-1}S_i\left(s_i-S_i'B_{M\infty}\right)\right]$$

$$=\sum_{i=-\infty}^{\infty}E\left[\left(s_0-S_0'B_{M\infty}\right) S_0'R_M^{-1}S_i\left(s_i-S_i'B_{M\infty}\right)\right]+o\left(\frac{M}{N}\right) . \quad (7)$$

Since $\{s_i\}$ is a Gaussian time sequence and $E\left[S_i\left(s_i-S_i'B_{M\infty}\right)\right]=0$, the series reduces to

$$\sum_{i=-\infty}^{\infty}\left\{E\left[\left(s_0-S_0'B_{M\infty}\right)\left(s_i-S_i'B_{M\infty}\right)\right] E\left[S_0'R_M^{-1}S_i\right]\right.$$

$$\left.+E\left[\left(s_0-S_0'B_{M\infty}\right)S_i'\right] R_M^{-1}E\left[S_0\left(s_i-S_i'B_{M\infty}\right)\right]\right\} . \quad (8)$$

Note that if

$$f_i(\lambda)=\sum_{-\infty}^{\infty}R_i(n)e^{in\lambda}$$

is the cross spectral density of arbitrary random processes, then

$$\sum_{n=-\infty}^{\infty} R_1(n)\, R_2(n) \;=\; \frac{1}{2\pi}\int_0^{2\pi} f_1(\lambda)\overline{f_2(\lambda)}\,d\lambda \quad .$$

Therefore Expression 8 can be put in the spectral form

$$\frac{1}{2\pi}\int_0^{2\pi} \left|1-e_M'(\lambda)B_{M\infty}\right|^2 p(\lambda)\, e_M^*(\lambda) R_M^{-1} e_M(\lambda) p(\lambda)\,d\lambda$$

$$+\frac{1}{2\pi}\int_{-\pi}^{\pi} (1-e_M^*(\lambda)B_{M\infty})\, e_M'(\lambda) p(\lambda) R_M^{-1} e_M(\lambda)\; (1-e_M^*(\lambda)B_{M\infty}) p(\lambda)\,d\lambda\,, \tag{9}$$

where $p(\lambda)$ is the process spectral density and $e_M(\lambda)$ is the M-column vector

$$e_M(\lambda) \;=\; \left\{e^{im\lambda},\ m=1,\cdots,M\right\}.$$

Here $(\ )'$ denotes transpose and $(\ )^*$ denotes complex conjugate transpose. Now R_M^{-1} can be expressed as the product of a triangular matrix with its transpose

$$R_M^{-1} \;=\; A_M' A_M \quad ,$$

where $A_M S_i$ is a random variable with uncorrelated components. Therefore,

$$A_M S_i \;=\; \left\{\frac{s_{i-k} - S_{k-1}' B_{(k-1)\infty};\ k=1,\cdots,M}{\sigma\,[k-1,\infty]}\right\}$$

and

$$A_M e_M(\lambda) \;=\; \left\{\frac{e^{ik\lambda}\ \left[1-e_{k-1}^*(\lambda)B_{(k-1)\infty}\right]}{\sigma\,[k-1,\infty]}\ ;\ k=1,\cdots,M\right\}$$

Expression 8 becomes

$$\sum_{k=0}^{M-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \left| 1-e_M'(\lambda) B_{M\infty} \right|^2 \frac{\left| 1-e_k^*(\lambda) B_{k\infty} \right|^2}{\sigma^2[k,\infty]} \right.$$

$$\left. + \left[1-e_M^*(\lambda) B_{M\infty}\right]^2 \frac{\left[1-e_k^*(\lambda) B_{k\infty}\right]^2}{\sigma^2[k,\infty]} e^{2i(k+1)\lambda} \right\} p^2(\lambda)\, d\lambda$$

But it is well known (e.g., Reference 6) that

$$\left| 1 - e_k'(\lambda)\, B_{k\infty} \right|^2 p(\lambda)$$

converges in measure to $\sigma^2[k, \infty]$. Hence, Expression 9 becomes

$$M \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \sigma^2[M,\infty] + e^{2i(M+1)\lambda} \left[\frac{1 - e_M^*(\lambda) B_{M\infty}}{1 - e_M'(\lambda) B_{M\infty}} \right] \right\} d\lambda + o(M)$$

$$= M\sigma^2[M,\infty] + o(M). \tag{10}$$

Substituting Equation 10 into Equation 7 and Equation 7 into Equation 6 results in

$$E\left[F^2(M,N)\right] = \sigma^2[M,\infty] \left(1 - \frac{M}{N}\right) + o\left(\frac{M}{N}\right) + \frac{o(M)}{N}. \tag{11}$$

MEAN SQUARE PREDICTION ERROR ESTIMATION

A comparison of Equations 5 and 11 immediately suggests the estimator,

$$\hat{\sigma}^2[M,N] = F^2(M,N) \frac{1 + (M/N)}{1 - (M/N)}. \tag{12}$$

The estimate of Equation 12 has the property that

$$E\ \hat{\sigma}^2[M, N] = \sigma^2[M, N] + o\left(\frac{M}{N}\right) + \frac{o(M)}{N} . \qquad (13)$$

The value of M for which Equation 12 is a minimum, denoted M_o, is an estimate of the value of M to be used in predicting s_o to minimize the error variance for the learning period N. The minimization is over the values of M up to some chosen maximum value M_{max}. The value of N is chosen so that $M_{max}/N << 1$ (typically 0.2-0.4). Then an estimate of the minimum variance $\sigma^2[M_{max}, \infty]$, is taken as

$$\hat{\sigma}^2[M_{max}, \infty] = F^2(M_{max}, N) \cdot \frac{1}{1 - (M_{max}/N)} .$$

Then N is adjusted to a value N_o such that

$$\frac{\hat{\sigma}^2\ [M,N]}{\hat{\sigma}^2[M_{max}, \infty]} \leq 1 + \epsilon,$$

where ϵ is an allowable deviation of the mean square prediction error from the optimum (typically about 0.2).

The analysis of the error term effect in Equation 13 is difficult. However, since it is the large changes in σ^2 [M,N] that are most important, for "sufficiently small" values of M_{max}/N the higher order effects are negligible. Unlike the fitting error itself, an estimate of σ^2 [M,N] has been produced which has a minimum. Another difficulty lies in determining the variance of the prediction error when this procedure is used. It should be noted that the calculation of this variance involves the determination of the probability as a function of N and M so that $\hat{\sigma}^2$[M,N] is a minimum. Since this is extremely complicated, computer simulations have been resorted to in an effort to verify the effectiveness of the procedure. Some results are given in the next section.

RESULTS

The prediction procedure described above has been extensively verified by computer simulation. Independent Gaussian random variables are generated and operated on by an appropriate digital filter to obtain data that has a given spectral density. The prediction procedure is repeated 10,000 times independently to give a 90 percent confidence interval with ± 2.33 percent of the true value of $\sigma^2[M_o, N_o]$. In general, the method has produced good results. The following typical results are presented for one spectral density. The spectral

density is chosen, as in Reference 4, to be bandlimited with a small sample-to-sample independent "noise" component

$$p(\lambda) = 4.996 \quad |\lambda| \leq 0.2\pi$$

$$= 0.001 \quad |\lambda| > 0.2\pi.$$

The memory and learning period are constrained by

$$M_{max} = 8$$

$$N \geq 20.$$

The values of M_o and N_o satisfy

$$\frac{\hat{\sigma}^2[M_o, N_o]}{\hat{\sigma}^2[M_{max}, \infty]} \leq 1.2;$$

N is changed by 10 at a time to satisfy this inequality starting with N = 20. The results obtained were

Prediction error variance over 10,000 samples = $\sigma_a^2 = 0.0120$,

Average learning period = $\bar{N}_o = 31.0$, and

Average memory = $\bar{M}_o = 5.95$.

The actual minimum variance for this spectral density is found to be

$$\sigma^2[8, \infty] = 0.00802.$$

Thus

$$\frac{\sigma_a^2}{\sigma^2[8, \infty]} = 1.50,$$

which is reasonably close to the desired maximum value of 1.2.

In Reference 4 the mean square prediction error for these data was a minimum for a memory of 4 with a fixed learning period of 30, the minimum being 0.0117 on the basis of computer simulations. Since this is close to σ_a^2, the procedure comes very close to the minimum for a learning period of 30. Therefore, it can be concluded that the procedure is reasonably effective for these data. Further computer simulations have borne out this conclusion more generally. For smaller values of ϵ and M_{max}/N, the desired value of $1 + \epsilon$ is naturally approached more closely.

SOME ADDITIONAL COMMENTS

It is recognized that, in the use of the foregoing procedure, it will not be detected if $\sigma^2[M,\infty]$ decreases significantly for $M > M_{max}$. However, since the procedure must be terminated somewhere, this risk will always exist.

Although in the absence of other information it seems most reasonable to weight those points closest to the value to be predicted, it is worth mentioning that the above analysis can be carried through for any linear predictor with M degrees of freedom. Thus an even more complicated procedure could be based on the minimum of the fitting error over all linear predictors with M degrees of freedom for various values of M.

An unknown factor is the effect of departures from the Gaussian assumptions as well as the effect of nonstationarities. Each possible factor has to be considered on its own merit.

REFERENCES

1. Balakrishnan, A. V., "An Adaptive Non-Linear Data Predictor," Proc. National Telemetering Conf. (May 23-25, 1962, Washington, D.C.), New York, N. Y.: The Inst. Radio Engineers, Paper 6-5, Vol. II, pp. 1-15.

2. Balakrishnan, A. V., "Effect of Linear and Non-Linear Signal Processing on Signal Statistics," J. Res. Natl. Bur. Std., 68D(9):953-965, September 1964.

3. Davisson, L. D., "A Theory of Adaptive Data Compression," Vol. 2 of Advances in Communication Systems, Ch. 5, A. V. Balakrishnan, Ed. Academic Press (to be published).

4. Davisson, L. D., "The Prediction Error of Stationary Gaussian Time Series of Unknown Covariance," IEEE Trans. on Information Theory, IT-11(4):527-532, October 1965.

5. Cramer, H., "Mathematical Methods of Statistics," Princeton, N.J.: Princeton Univ. Press, 1946, p. 368.

6. Grenander, U., and Szego, G., "Toeplitz Forms and Their Applications," Berkeley, Calif.: Univ. of Calif. Press, 1946.

3. RESULTS OF ADAPTIVE PREDICTOR STUDIES, PART 1

R. L. Kutz

Goddard Space Flight Center
Greenbelt, Maryland

The data used for this study are taken from ten Tiros cloud pictures. The Tiros video data are normally stored on analog tape at the ground station. To be useful for digital processing, the analog data are sampled, quantized to six bits, and stored on digital magnetic tape with one sample per tape character. A digital computer is used to remove the line synchronization pulses that are necessary only for the analog picture display equipment. The fifth picture of the series is used in the following discussion as a typical example of the ten-picture group. An example of an analog picture produced by the Tiros kinescope ground station display equipment is shown in Figure 3-1 and the digital version of the same picture after the first computer processing is shown in Figure 3-2. All digital pictures are displayed using the Stromberg Carlson 4020 high-speed microfilm recorder. In order to make an evaluation of the processing effect on the picture, it is necessary to redisplay cloud pictures which have been subjected to data compression.

The first data compression technique to be discussed is that of linear prediction (Reference 1). In this type of prediction a weighting of past data samples is used to produce an estimate of the sample to be predicted. The first point of interest is to determine the memory M, i.e., how many past samples need to be weighted to get the best estimate of the following data sample in the sequence. This was done by using L (the learning period) data samples to produce a set of weights which are used to predict the following L/2 samples with an error threshold $T_1 = 0$. Figure 3-3 indicates that M = 3 produces the minimum rms prediction error for L = 20. Figure 3-4 shows the effect on the rms learning and prediction errors of changing L with M fixed at 3.

It is undesirable to use zero error data samples to produce a set of weights because of the implication that error-free data samples are available at the receiver. Error-free data samples arrive at the receiver in two ways — either they are predicted with no error or they are transmitted through the error-free channel. For the data used here, it was found that L of the learning samples frequently had to be transmitted. The savings to be achieved through data compression are negligible if many sets of weights are required, as is frequently the case.

Figure 3-1. Analog Picture Produced by the TIROS Kinescope Ground Station Display Equipment

As explained in Reference 1, half of the learning period was made up of transmitted elements a priori and the other half was made up of processed data having an error of less than T_1. A rule requiring no special coding is used to determine when a new set of weights is to be generated. Since the weights are to be identical at the transmitter and the receiver, they must be derived from an identical learning set of points. This rule involves the testing of a second threshold T_2 against the mean square prediction error on transmitted elements every time $L/2$ samples have been processed. Threshold T_2 is chosen by compressing the same data for several values of T_2, with the other parameters fixed, and picking the value of T_2 which maximizes the compression. A value of T_2 which is too small will constantly demand new weights and, therefore, much of the data will be transmitted without any attempt at prediction. On the other hand, a value of T_2 which is too large will cause few sets of weights to be produced, hence a set of weights may be often used over portions of the data where it performs poorly.

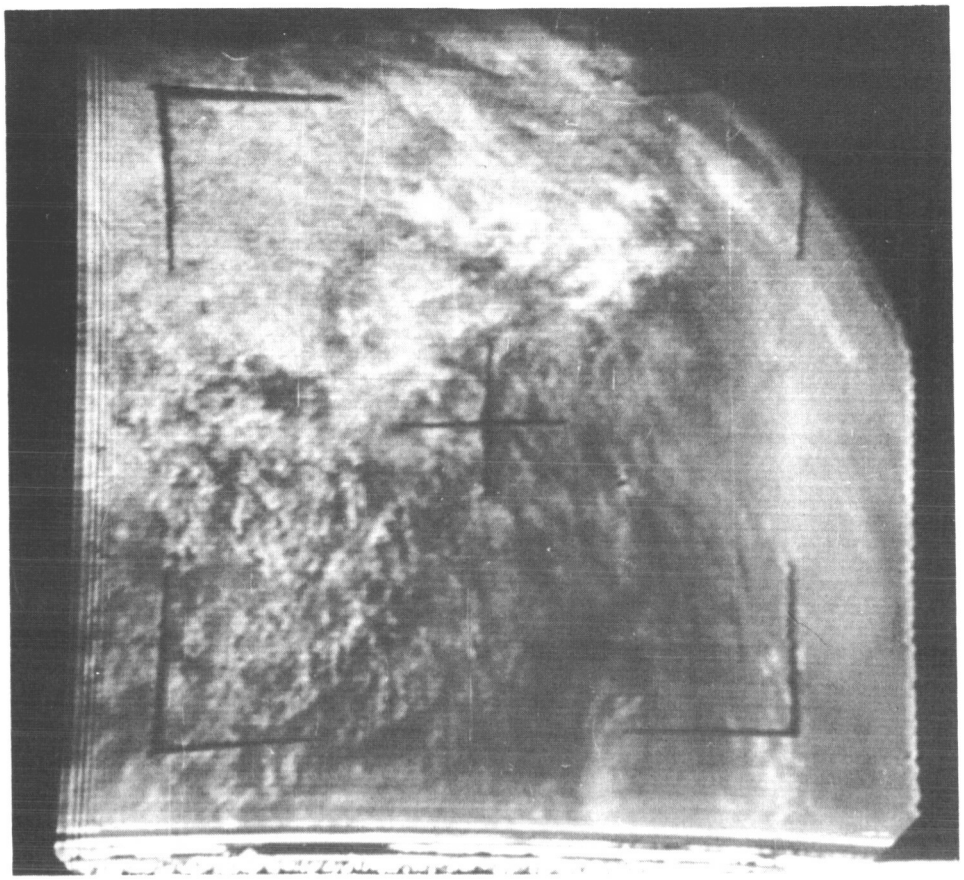


Figure 3-1. Analog Picture Produced by the TIROS Kinescope Ground Station Display Equipment

As explained in Reference 1, half of the learning period was made up of transmitted elements a priori and the other half was made up of processed data having an error of less than T_1. A rule requiring no special coding is used to determine when a new set of weights is to be generated. Since the weights are to be identical at the transmitter and the receiver, they must be derived from an identical learning set of points. This rule involves the testing of a second threshold T_2 against the mean square prediction error on transmitted elements every time $L/2$ samples have been processed. Threshold T_2 is chosen by compressing the same data for several values of T_2, with the other parameters fixed, and picking the value of T_2 which maximizes the compression. A value of T_2 which is too small will constantly demand new weights and, therefore, much of the data will be transmitted without any attempt at prediction. On the other hand, a value of T_2 which is too large will cause few sets of weights to be produced, hence a set of weights may be often used over portions of the data where it performs poorly.

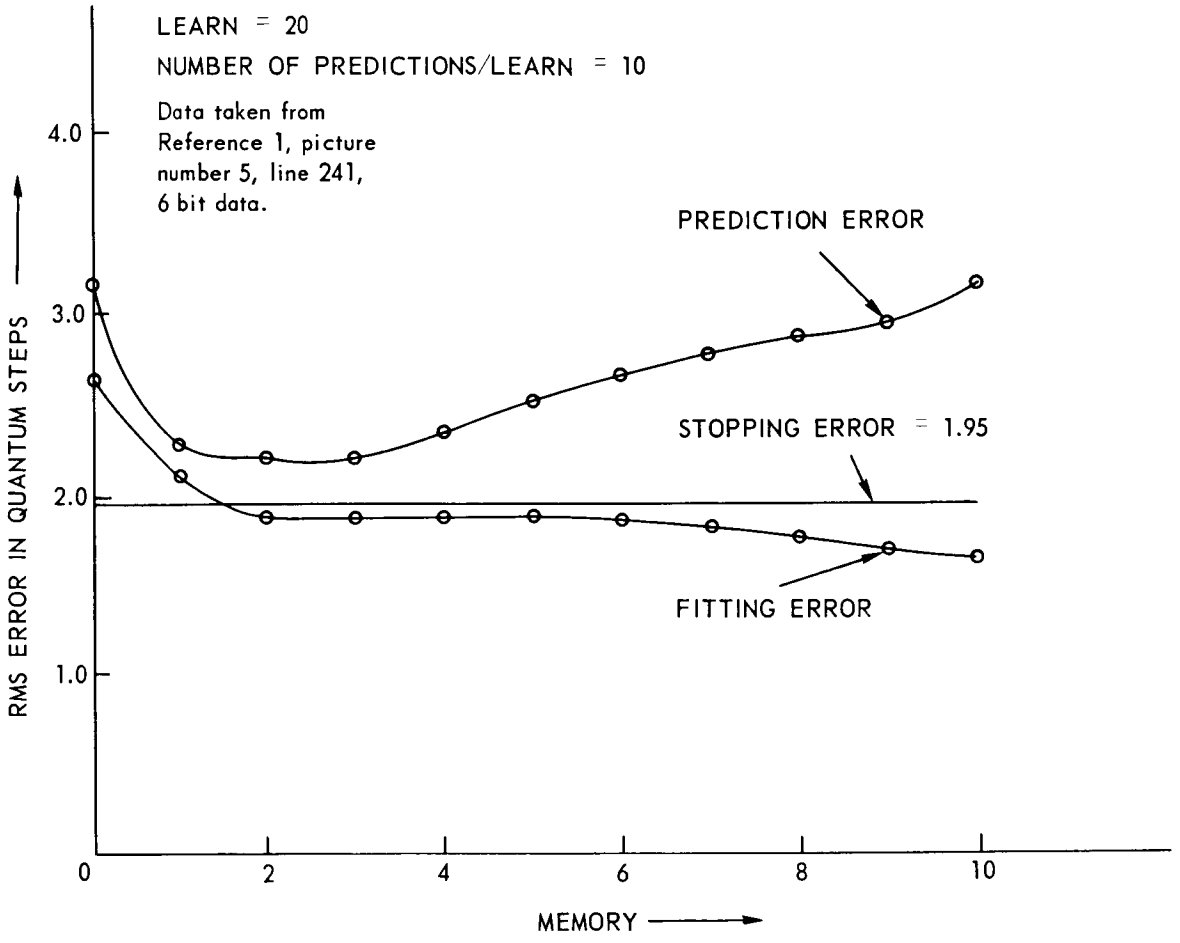


Figure 3-3. Prediction and Learning Errors Versus Memory

term, 10 weights are required for D = 2. The results are shown in Table 3-1 for comparative values of M and L. Here the learning period is too short for 10 weights, which would account for the poor results.

The method used to find the weights which would minimize the mean square fitting error over the learning period involves the solution of simultaneous equations. Both the "steepest descent" (SD) and "conjugate gradient" (CG) methods of solution for simultaneous equations have been tried. The CG method converges in fewer iterations than does the SD method, but the SD method has fewer matrix operations per iteration than does the CG method and the fitting error is produced as part of each cycle, which is of interest in controlling the quality of fit.

Table 3-1

COMPARISON BETWEEN TWO CLASSES OF PREDICTORS
AND BETWEEN THREE COMPRESSION TECHNIQUES

Prediction Method	Learn L	Memory M	Parameter D	Threshold T_1	Element Compression	Bit Compression (Method 2)	Bit Compression (Entropy Code)
Linear (Method 1)	1*	0*	1*	2.0*	5.42*	2.47*	3.10*
	20	3	1	2.5	5.09	2.41	2.94
	20**	3**	1**	2.5**	2.10**	Not available**	1.48**
	20	3	2	2.5	3.30	1.75	2.08
	10	1	1	2.5	5.06	2.57	2.93
	14	1	2	2.5	4.95	2.50	2.88
Markov (Method 2)	Learn restarted every 16 lines	1	—	2.0	5.95	Not available	3.34

* Zero-order hold.
**Old method, learn ½ on predicted elements and ½ on unprocessed (0 error) data.

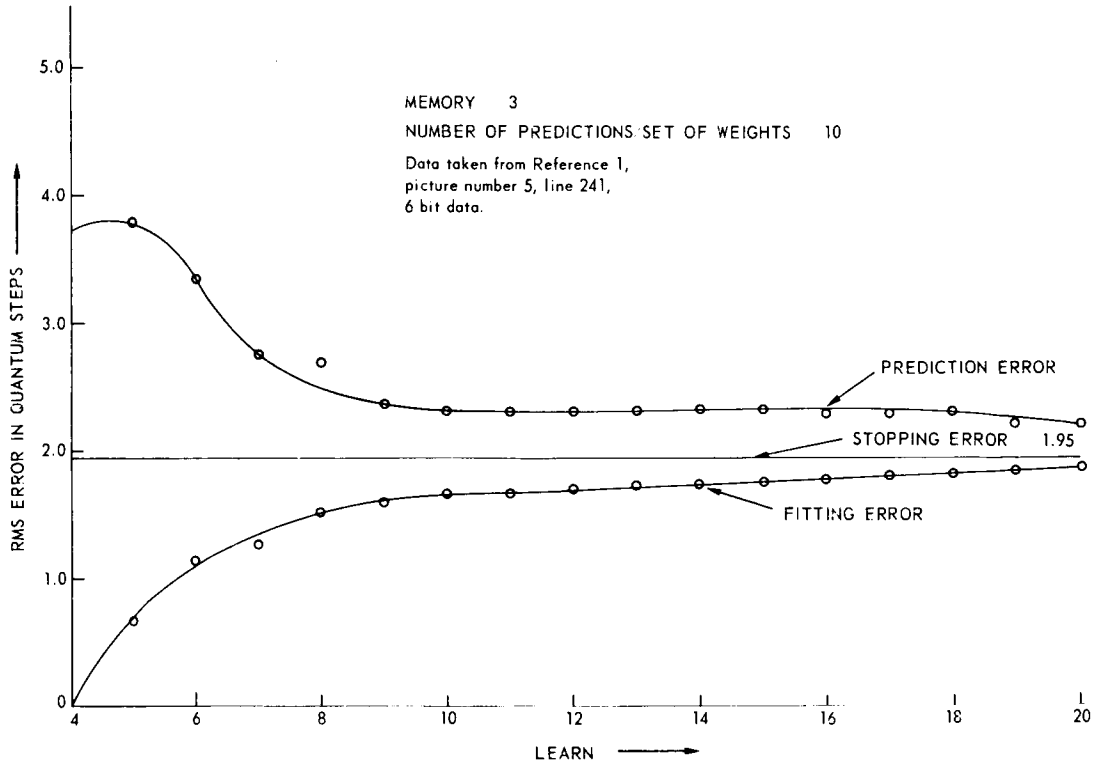


Figure 3-4. Prediction and Learning Error Versus Learn

In general, the compression ratio is quite sensitive to the quality of the fit over the learning period. The closest fit over the learning period does not always produce the highest compression ratio, depending on the size of M and L. However, for the values of M and L considered here and for D = 1, a minimum mean square fit gives the best results. In general, the constant term (Reference 1) should not be important for a minimal fit solution. This is not found to be the case when the SD solution is used as shown in Figure 3-5. The constant affects the magnitude of the terms and hence emphasizes the truncation error. An IBM 7094 computer is used for this simulation with eight decimal digits of accuracy.

The linear predictor with M = 0 and L = 1, which is the zero-order hold, is used as a standard for the comparison of various data compression techniques. Table 3-1 shows a comparison between two classes of predictors, linear and Markov, with five variations of the linear predictor including the zero-order

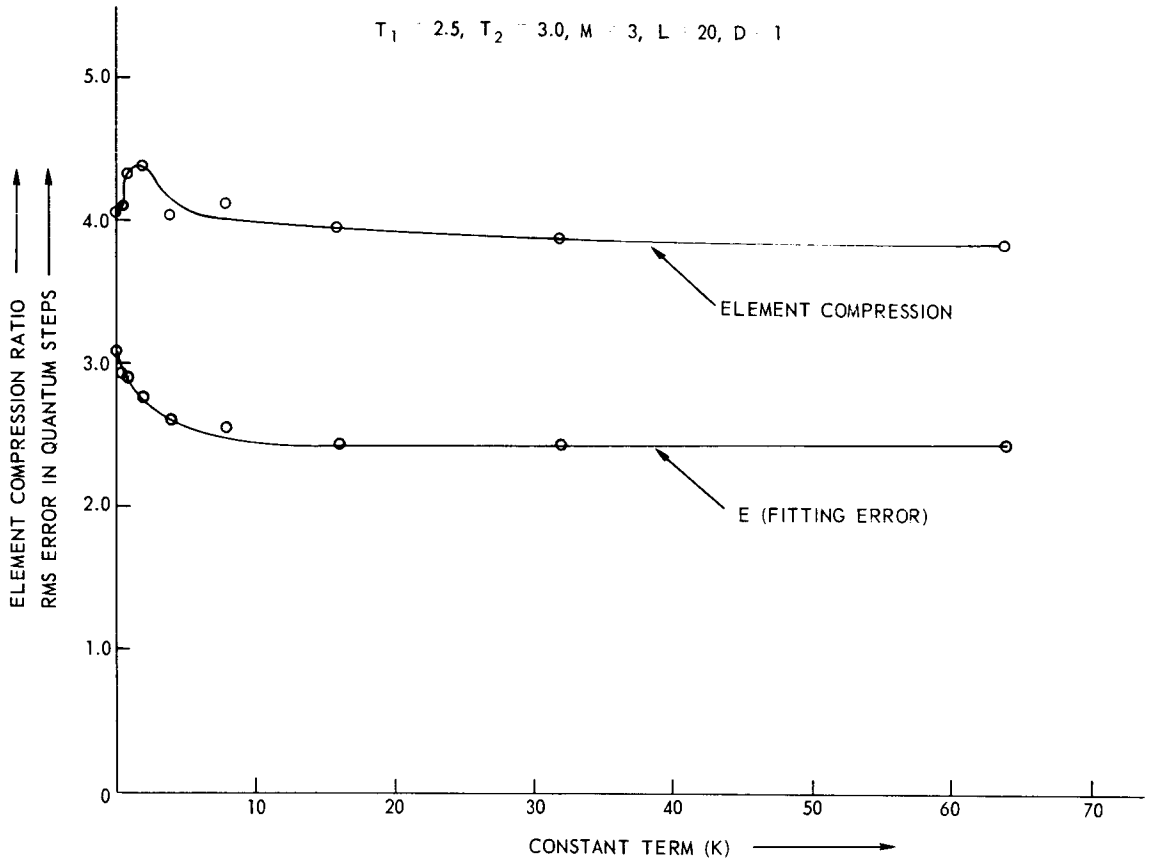


Figure 3-5. Element Compression and E (Fitting Error) Versus Constant Term (K)

hold. There are three compression ratios shown in Table 3-1. The first is the element compression ratio defined by

$$\text{Element compression} = \frac{\text{(Number of TV elements in original digital picture)}}{\text{(Number of TV elements not predicted, i.e., the number of transmitted elements)}},$$

and (q = 1/Element Compression) is the estimate of the probability of transmission, as if such a probability existed. The second and third compression ratios give the bandwidth reduction, the equivalent transmission time, or the power reduction achievable in a noiseless channel. The second compression ratio is produced using run-length coding. Here, run-length coding (Method 2) uses an indicator bit with each transmitted data word to indicate whether the remaining word bits are a TV intensity element or a portion of a run-length word. The

third compression ratio is calculated from the entropy equation using q as the probability of transmission and the number of bits used to represent each TV element's intensity N = 5 where

$$\text{Bit compression (entropy code)} = \frac{N}{qN - q \log_2 q - (1-q)\log_2(1-q).}$$

The usual assumption is made in the above equation, i.e., that all intensities are equally likely after prediction, which is the worst case. In Reference 1, another type of run-length coding is discussed and is called Method 1 coding. Using Method 1 coding, K bits are added to each transmitted TV element to allow for run lengths up to 2^K elements. However, Method 1 coding has the disadvantage that a data expansion from N bits per element to (N + K) bits per element results if no compression is possible. The second compression figure in Table 3-1, labeled bit compression (method 2), also results in a slight expansion where no compression is possible; however, it is small i.e., from N bits per sample to (N + 1) bits per sample.

Table 3-1 shows that of the linear predictors the zero-order hold performs best. The results shown for the Markov predictor, which are discussed in detail in part II of this study*, are more favorable than those of the linear predictors. A Markov predictor with a memory of 1 is shown in Table 3-1 because no cases of memories greater than 2 have been programmed for five-bit data because of computer storage size and because a memory of 1 is easier to discuss and graph. Learning over 16 lines, as with the Markov predictor, implies a slower adaption of the predictor than with the linear predictor, which also has a memory of 1 but learns over 10 to 14 elements. Table 3-1 shows the linear predictor with a memory of 1 for D = 1 and D = 2. As before, the D = 2 case is not so good as the D = 1 case. It should be noted that the linear predictor with a memory of 1 performs nearly as well as the linear predictor with a memory of 3.

Figure 3-6 shows a comparison between the linear and Markov predictors for a memory of 1 and a learning period which uses all of the picture elements. To get a prediction using Figure 3-6, the previous element value is placed on the abscissa and, using the desired curve, the predicted value is read on the ordinate. There is a noticeable deviation between the two methods for those intensities greater than 25 as well as for those intensities less than 10. This difference in the prediction curves of Figure 3-6 indicates a difficulty in the linear predictor which could account for its poor performance. That is, one may have to use a

*Paper 6 of this conference. Prepared by J. A. Sciulli, Goddard Space Flight Center.

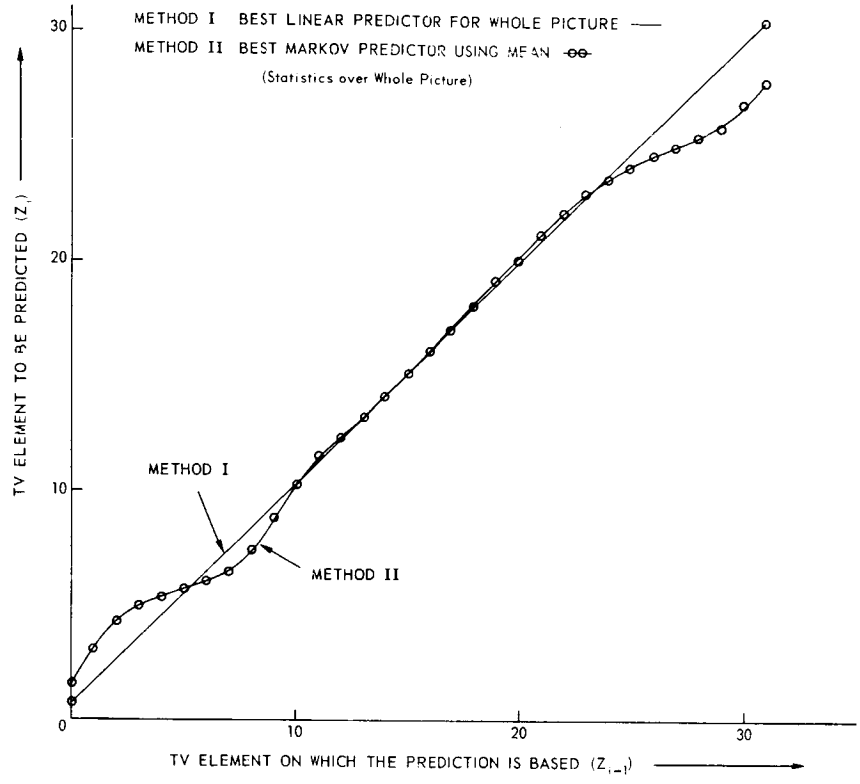


Figure 3-6. TV Element to be Predicted Versus TV Element on Which the Prediction is Based.

prediction curve that is not optimal, in the mean square error sense, for some of the data intensities. A method of combining the rapid learning of the linear predictor with the optimality of the Markov predictor at each intensity amplitude is given in Reference 2 as Method 3.

It should be mentioned that rather than using the conditional expectation (i.e., conditional mean as reported herein) to predict one could use the conditional mode. The conditional mode has been used in Reference 3 by L. D. Davisson, with somewhat better results than those obtained by use of the conditional mean.

Figures 3-7 through 3-10 show the quality of TV pictures which have been compressed, transmitted through an error-free communication channel, and displayed after expansion. The compression technique used is the linear prediction with M = 3, L = 20, and D = 1 for T_1 = 2.5 (Figure 3-7), T_1 = 3.5 (Figure 3-8),

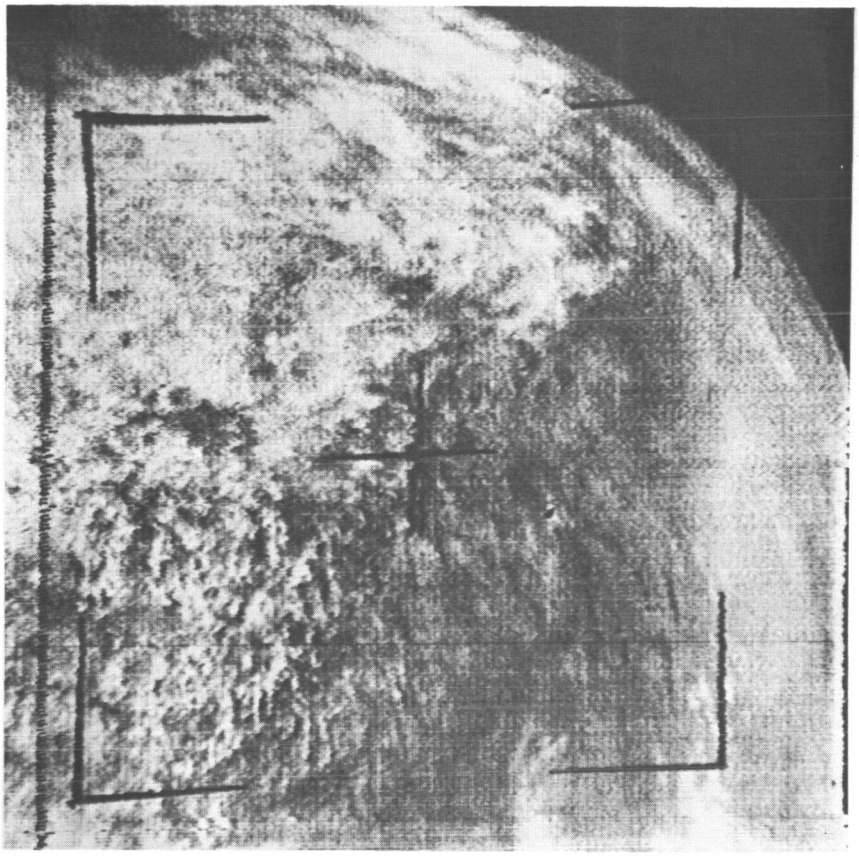


Figure 3-2. Digital Version of Figure 3-1 After the First Computer Processing

In order to simplify the selection of T_2 and avoid the transmission of too many data samples, a different compromise is made. The learning set is made up entirely of processed data which may have sample errors as large as T_1. As might be expected, a small loss in compression is encountered for fixed values of T_1 and T_2 which correspond to the optimum value for the previous method of learning. However, now the only cost due to making T_2 smaller is the increased processing time for the computer. As shown in Table 3-1, a small value of T_2 results in an improvement in compression over the previous method.

Another possible way to increase the compression is by using nonlinear prediction. Here a parameter D is used to indicate that all samples of the memory are combined in such a way that each uniquely indexed product of samples, i.e., D, D -1, etc. down to the individual samples themselves, are weighted to get each estimate. For example, with a memory of 3 and a constant

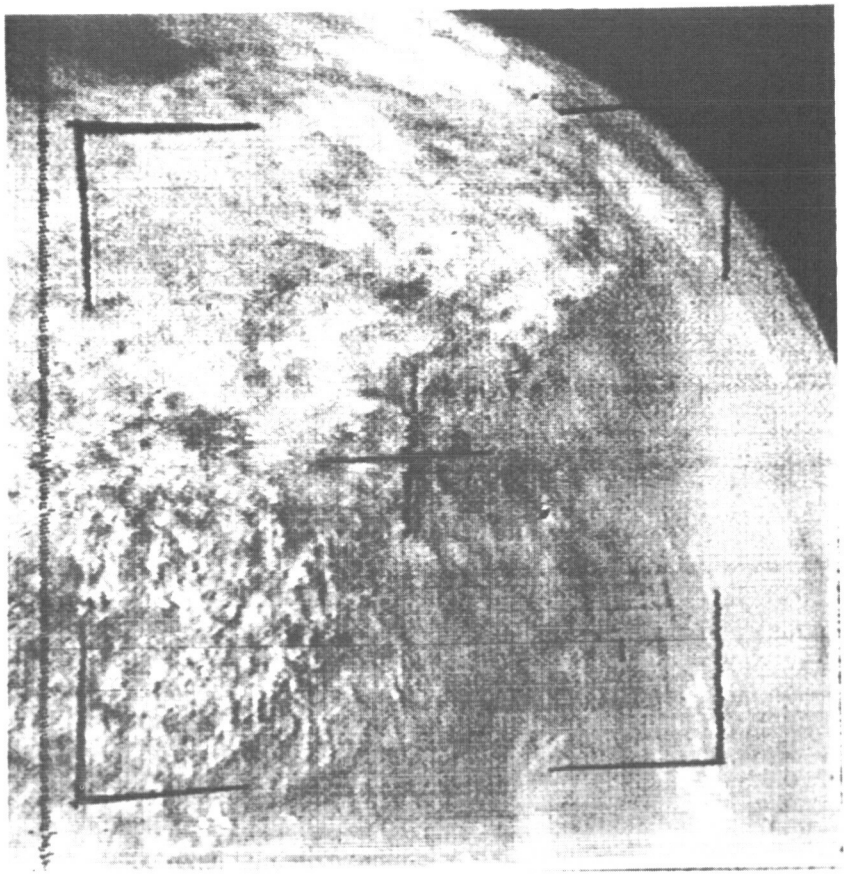


Figure 3-7. TV Picture After Compression, Transmission, and Expansion where $T_1 = 2.5$

$T_1 = 5.5$ (Figure 3-9), and $T_1 = 7.5$ (Figure 3-10). Those pictures with $T_i > 3.5$ would be unuseable.

CONCLUSION

When using linear prediction, higher compression ratios can be achieved by learning from data for which prediction has been attempted. A memory of 3 gives slightly better results than a memory of 1 using linear prediction, but the difference is negligible. The zero-order hold is the best of the linear predictors. If nonlinear terms are introduced, as with $D = 2$, the compression ratios are lower than with $D = 1$ using linear prediction. The two methods mentioned for solving the minimum mean square error weights are "Conjugate Gradients" and "Steepest Descents," of these, the latter is used most. Table 3-1 shows that a more efficient code could gain only 1 db at most for the data being considered. Table 3-1 is also used to make a comparison between the linear and Markov

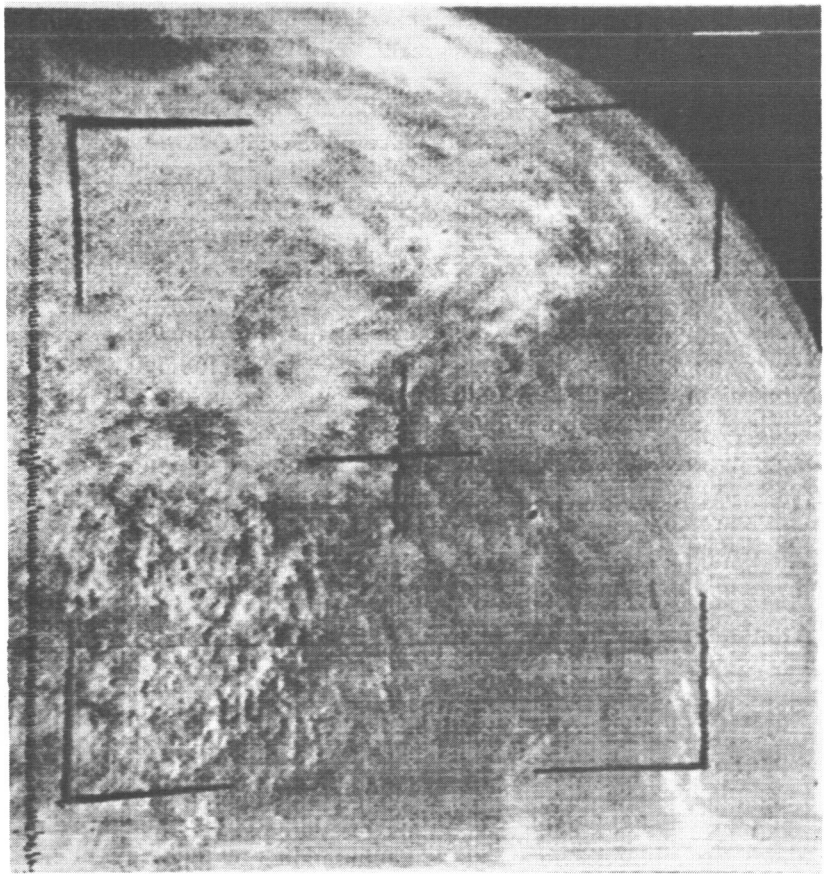


Figure 3-8. TV Picture After Compression, Transmission, and Expansion where $T_1 = 3.5$

predictors. The Markov predictor performs better than the linear predictors probably because of its optimality for all data amplitudes. It is suggested that Balakrishnan's Method 3 be used to circumvent the problems encountered with the other two methods.

REFERENCES

1. Balakrishnan, A. V., Kutz, R. L., and Stampfl, R. A., "Adaptive Data Compression for Video Signals," NASA Technical Note D-3395, April 1966.

2. Balakrishnan, A. V., "An Adaptive Nonlinear Data Predictor," Proc. National Telemetering Conf. (May 23-25, 1962, Washington, D.C.), New York, N. Y.: The Inst. of Radio Engineers, Vol. II, pp. 1-15.

Figure 3-9. TV Picture After Compression, Transmission, and Expansion where $T_1 = 5.5$

3. Davisson, L. D., "Data Compression and its Application to Video Signals," Analysis of Space Data and Measurement of Space Environments, Final Rept. GSFC Summer Workshop Program, GSFC Document X-100-65-307, June 15 to September 15, 1965, pp. A-35-A-55.

4. DEMONSTRATION OF A QUANTILE SYSTEM FOR COMPRESSION OF DATA FROM DEEP SPACE PROBES

T. O. Anderson, I. Eisenberger, W. A. Lushbaugh, and E. C. Posner

Jet Propulsion Laboratory, CIT
Pasadena, California

THEORETICAL BACKGROUND OF THE QUANTILER

In circumstances such that the locality of experiments does not coincide with the locality of the statistical computations performed on the experimental data, the sample data must be transmitted through a communications channel from the point of origin to the processing center. If many experiments are being performed simultaneously and the same communication channel is used to transmit the sample values of each experiment, there is a limit to the total number of observations that can be sent in a given time. This limits the number of experiments that can be performed for given sample sizes. One way of effecting "data compression" is by transmitting a small number of sample quantiles (instead of all of the sample values) and extracting the desired statistical information. The following is a brief discussion of the use of sample quantiles for this purpose.

Let $x_1, x_2, \ldots, x_n$ be n independent sample values taken from a population with a density function $g(x)$ and a distribution function $G(x)$. If $g(x)$ is continuous, then ζ_p is said to be the (population) quantile of order p, or the p^{th} quantile of the distribution, if ζ_p is the (unique) solution of the equation $G(\zeta) = p$. Equivalently,

$$p = \int_{-\infty}^{\zeta_p} g(x)\, dx, \quad (0 < p < 1).$$

Similarly, a sample quantile of order p, Z_p, can be defined by arranging the observations in ascending order of magnitude,

$$x_{(1)} \leqq x_{(2)} \leqq, \ldots, \leq x_{(n)}.$$

Then

$$Z_p = x_{[np]} ,$$

where $[np]$ denotes the greatest integer $\leq np$. This means that 100 percent of the sample values are smaller than Z_p. For example, $Z_{1/2}$ is well-known as the median of the sample distribution.

Now the sample quantiles, being random quantities, have probability distributions of their own. Moreover, subject to the condition that the (original) parent density function possesses a continuous derivative in some neighborhood of each quantile value being considered, the joint distribution of any (fixed) number of quantiles has the useful property of approaching the multivariate normal distribution as the sample size $n \to \infty$ (Reference 1). The mean and variance of the limiting distribution of Z_p are given in that reference by

$$E(Z_p) = \zeta_p$$

and

$$\mathrm{Var}(Z_p) = \frac{p(1-p)}{n g^2(\zeta_p)} .$$

The correlation ρ_{12} between Z_{p_1} and Z_{p_2}, for $p_1 < p_2$ is given by

$$\rho_{12} = \left[\frac{p_1(1-p_2)}{p_2(1-p_1)}\right]^{1/2} .$$

Thus, for a sufficiently large sample size, if the limiting normal distribution of the sample quantiles is assumed when statistical analyses are based on sample quantile values, the error involved in making the normality assumption will be small.

Estimates of the parameters of a normal population have been obtained by using quantiles. A measure of the reliability of such an estimate is its variance. The relative reliability, called the estimate efficiency, is then defined to be the ratio of the variance of the best estimate using the entire sample to that of the estimate using quantiles.

Another type of statistical information obtained by using quantiles is concerned with tests of simple hypotheses. Here a measure of reliability is the power of the test, that is, the probability of rejecting the assumed, or null, hypothesis when it is false. The relative reliability, also called the efficiency of the test, is defined in this context to be the ratio of the power of the test using quantiles to that of the best test using all of the sample values. The task of determining the efficiencies of the estimators and tests is simplified considerably by the assumption of the normality of the quantiles.

Data compression is obtained in the following manner. The use of a small number of quantiles to obtain statistical information means that only a few observations need to be transmitted instead of all of the sample values. The result is a high "data compression ratio." This compression ratio would be of little value, however, if the relative reliability of the information acquired in this manner were proportional to the ratio of the number of quantiles used to the total number of observations. The relative reliability, as shown, is not proportional to this ratio, but depends only upon the number of quantiles used and how they are chosen. This accounts for the advocacy of the use of quantiles for data compression.

The primary function then of a "quantiler" is to choose a number of quantiles of specified orders from a set of samples and transmit these values to Earth, where the desired information can then be extracted. Since, for a fixed number of quantiles, the relative reliability of a given type of statistical information is critically dependent upon the choice of the quantiles used, it is natural that one should specify the orders of the quantiles which maximize the relative reliability. However, there are certain restrictions in making this optimal choice.

First of all, the orders of the quantiles might have to be specified in advance and remain fixed throughout the flight. Second, the orders of the quantiles which maximize the relative reliability of one type of information are not necessarily those which maximize the relative reliability of another type. Hence, in order to use the same set of quantiles for all types of statistical information that may be desired, a compromise must be made in the choice of their orders. A brief theoretical discussion will therefore be given concerning the optimal choice of quantiles for estimating the mean μ and standard deviation σ of a normal parent population. Details are found in Reference 2.

If it is assumed that both the mean μ and standard deviation σ are unknown, an estimator of μ using $k \geq 2$ quantiles of distinct arbitrary orders of the form

$$\hat{\mu} = \sum_{i=1}^{k} C_i Z_i$$

is unbiased (has expected value μ) if the C_i are determined such that

$$\sum_{i=1}^{k} C_i = 1$$

and

$$\sum_{i=1}^{k} C_i \zeta_i^* = 0,$$

where each Z_i is of order p_i and ζ_i^* is the population quantile of order p_i of the standard normal distribution. This occurs, because when the expected value of $\hat{\mu}$ is used one has

$$E(\hat{\mu}) = \sum_{i=1}^{k} C_i \zeta_i = \sum_{i=1}^{k} C_i (\mu + \sigma \zeta_i^*) = \mu + \sigma \sum_{i=1}^{k} C_i \zeta_i^* = \mu.$$

It is obvious that for $k = 2$, unless Z_1 and Z_2 are symmetric quantiles (that is, unless $p_2 = 1 - p_1$, implying that $\zeta_1^* + \zeta_2^* = 0$), $\hat{\mu}$ will not be unbiased for any choice of C_1 and C_2. Moreover, for any even k, it has been shown that, in order for $\hat{\mu}$ to have minimum variance, pairs of symmetric quantiles should be used (Reference 3). Under this restriction, it is easily seen that the estimate of μ in the form

$$\hat{\mu} = \sum_{i=1}^{k/2} a_i (Z_i + Z_{k-i+1}), \qquad p_i + p_{k-i+1} = 1$$

will be unbiased if $a_1 > 0$, $i = 1, 2, \ldots, k/2$, and

$$\sum_{i=1}^{k/2} a_i = 1/2.$$

By the method of maximum likelihood, the coefficients a_i can be determined to provide an unbiased estimate of μ with minimum variance for $k/2$ arbitrarily chosen pairs of symmetric quantiles. Finally, by varying the p_i in the expression for the variance of $\hat{\mu}$, the orders of the particular set of $k/2$ pairs of symmetric quantiles, as well as the coefficients a_i for which $\mathrm{Var}(\hat{\mu})$ is a minimum, can be determined.

By using the same technique, unbiased estimators of σ from $k/2$ pairs of symmetric quantiles can be obtained; these estimators are of the form

$$\hat{\sigma} = \sum_{i=1}^{k/2} \frac{\beta_i}{\zeta^*_{k-i+1}} (Z_{k-i+1} - Z_i)$$

where $\beta_i > 0$ and

$$\sum_{i=1}^{k/2} \beta_i = 1/2.$$

For this case,

$$E(\hat{\sigma}) = \sum_{i=1}^{k/2} \frac{\beta_i}{\zeta^*_{k-i+1}} \left(\mu + \sigma\zeta^*_{k-i+1} - \mu - \sigma\zeta^*_i\right) = 2\sigma \sum_{i=1}^{k/2} \beta_i = \sigma,$$

since $\zeta^*_i = -\zeta^*_{k-i+1}$.

By again varying the p_i's in the expression for the variance of the maximum-likelihood estimator of σ, the values of the p_i and β_i that minimize $\mathrm{Var}(\hat{\sigma})$ can be determined. Only symmetric quantiles ($p_i + p_{k-i+1} = 1$) have been used to estimate σ (it has not yet been proved, but only conjectured, that this is the optimum procedure to adopt). It can be said at present about the $\hat{\sigma}$ obtained in this manner, that this $\hat{\sigma}$ is the maximum-likelihood minimum-variance unbiased estimator of σ using $k/2$ pairs of symmetric quantiles.

Further mathematical details are given in Reference 2, where estimators of μ for $k = 1, 2, 3, 4, 5, \ldots, 20$ and estimators of σ for $k = 2, 4, 6, \ldots, 20$ are also given, having been derived by the method described above. The efficiency of $\hat{\mu}$, as defined previously, is shown in Reference 2 to increase from

0.637 for k = 1 to 0.994 for k = 20. The efficiency of $\hat{\sigma}$ was shown to increase from 0.612 for k = 2 to 0.984 for k = 20.

Since μ is assumed to be unknown, an estimator for σ cannot be obtained using one quantile. The median is the only possible estimator of μ using one quantile if σ is unknown, since $\zeta^* \neq 0$ for values of p other than p = 0.5. In fact, the median is the minimum-variance estimator of μ, even if σ is known. If μ is known, however, an estimator of σ can be obtained using one quantile. Both Z(0.9424), the quantile of order 0.9424, and Z(0.0576), the quantile of order 0.0576, provide unbiased minimum-variance estimators, which are given by

$$\hat{\sigma}_1 = 0.635\left[Z(0.9424)\right] - \mu$$

and

$$\hat{\sigma}_2 = -0.635\left[Z(0.0576)\right] - \mu \ .$$

The efficiency of both estimators is 0.304.

The attempt to achieve a high data compression ratio by the use of quantiles is rewarded by the fact that the best estimators of μ and σ, using as few as four quantiles, attain reasonably high efficiencies. With four quantiles, $\mathrm{Eff}(\hat{\mu}) = 0.920$, $\mathrm{Eff}(\hat{\sigma}) = 0.824$, the estimators are given by

$$\hat{\mu} = 0.192\left[Z(0.1068) + Z(0.8932)\right] + 0.308\left[Z(0.3512) + Z(0.6488)\right]$$

and

$$\hat{\sigma} = 0.116\left[Z(0.9770) - Z(0.0230)\right] + 0.236\left[Z(0.8729) - Z(0.1271)\right] ,$$

as shown in Reference 2.

It can be seen, however, that the orders of the optimum quantiles for estimating μ are quite different from those for estimating σ, and similar disparities are apparent for all values of k. It is true that, for small deviations from the optimum values, the loss in efficiency of either estimator will be insignificant; this fortunate occurrence is attributed to the flatness of the surface representing $\mathrm{Var}(\mu)$ and $\mathrm{Var}(\sigma)$ about the optimum quantile values. Reference 2 discusses this flatness in more detail.

A more or less random choice of quantiles, however, can result in a serious reduction in efficiency. For example, if the optimum four quantiles for estimating σ are used to estimate μ, the efficiency of $\hat{\mu}$ is reduced to 0.732, a loss in efficiency of 25.7 percent. And, if the optimum four quantiles for estimating μ are used to estimate σ, the efficiency of σ is reduced to 0.665, a loss of 19.2 percent. The problem, then, becomes one of establishing a criterion of optimality on the basis of which suboptimum sets of k quantiles could be determined that would provide unbiased estimators of both μ and σ, using the same quantiles.

The decision was made to use as this criterion the linear combination

$$\text{Var}(\mu) + b\,\text{Var}(\sigma), \qquad b = 1, 2, 3, \ldots .$$

The set of k/2 pairs of symmetric quantiles which minimized this linear combination was then defined as being suboptimum. For b = 1, 2, 3 and for k = 2, 4, 6, $\cdots$, 20, the suboptimum sets of quantiles, in the above sense, were determined, the estimators constructed, and the efficiencies computed. These results are given in Reference 2.

The results for k = 4 are as follows:

For b = 1,

$$\hat{\mu} = 0.141\left[Z(0.0668) + Z(0.9332)\right] + 0.359\left[Z(0.2912) + Z(0.7088)\right],$$

$$\hat{\sigma} = 0.258\left[Z(0.9332) - Z(0.0668)\right] + 0.205\left[Z(0.7088) - Z(0.2912)\right],$$

$$\text{Eff}(\hat{\mu}) = 0.908, \text{ and } \text{Eff}(\hat{\sigma}) = 0.735.$$

for b = 2,

$$\hat{\mu} = 0.106\left[Z(0.0434) + Z(0.9566)\right] + 0.394\left[Z(0.2381) + Z(0.7619)\right],$$

$$\hat{\sigma} = 0.196\left[Z(0.9566) - Z(0.0434)\right] + 0.232\left[Z(0.7619) - Z(0.2381)\right],$$

$$\text{Eff}(\hat{\mu}) = 0.876, \text{ and } \text{Eff}(\hat{\sigma}) = 0.779.$$

For b = 3,

$$\hat{\mu} = 0.097\left[Z(0.0389) + Z(0.9611)\right] + 0.403\left[Z(0.2160) + Z(0.7840)\right],$$

$$\hat{\sigma} = 0.179\left[Z(0.9611) - Z(0.0389)\right] + 0.235\left[Z(0.7890) - Z(0.2160)\right],$$

$$\text{Eff}(\hat{\mu}) = 0.857, \text{ and } \text{Eff}(\hat{\sigma}) = 0.792.$$

It is readily seen that as b increases $\text{Eff}(\hat{\mu})$ decreases and $\text{Eff}(\hat{\sigma})$ increases. Since larger values of b mean that greater weight is being given to $\text{Var}(\hat{\sigma})$, this is not a surprising result. The point is, however, that the value of b should logically be determined by the two estimators. If $\hat{\sigma}$ is the important consideration, increasing the value of b gives different suboptimum sets of quantiles for which the efficiency of $\hat{\sigma}$ is improved. If $\hat{\mu}$ is of paramount importance, b should be small. It should be noted, perhaps, that five choices of b are presently available (including b = 0 and, in effect, b = ∞).

One may in some cases prefer to suffer a loss in efficiency rather than use optimum minimum-variance quantiles. This occurs because, when k increases, the two optimum extreme quantile values, those of order p_1 and $p_k = 1 - p_1$, move farther out on the tails of the distribution. Although from a theoretical viewpoint this fact is of little consequence, from practical considerations there are two major objections to this behavior of p_1 and p_k. First, since n is never infinite, the true distributions of the sample quantiles are only approximately normal and, more important, the deviation from normality becomes more pronounced the farther the quantiles move out on the tails of the distribution. Second, the "normal" distributions that one encounters in practical situations are very often only approximately normal, with deviations from normality greater out on the tails than toward the center of the distribution. Thus, it is important on both counts to investigate the effect on the efficiency of the estimators when optimum and suboptimum estimates of μ and σ are obtained and when p_1 is restricted to be not less than some specified value. If the loss in efficiency is not excessive, it may well prove advantageous to adopt the cautious policy of restricting the value of p_1, and thus avoid or limit a bias in the estimates of μ and σ. This loss can be due either to a sufficiently large deviation from the assumed normality of the extreme quantile distribution or to the erratic behavior out on the tail of an approximately normal parent distribution.

Optimum and suboptimum estimators of μ and σ were constructed when p_1 was restricted to be not less than 0.01 and again for p_1 not less than 0.025. The

results were quite satisfactory in that losses in efficiency were small. Although the restrictions on p_1 affect the efficiencies of $\hat{\sigma}$ to a greater extent than those of $\hat{\mu}$, efficiencies greater than 0.90 can be achieved for suboptimum estimators of σ for $k \geq 10$ when p_1 is restricted to not less than 0.025. In fact, for this case very little is gained by using more than 10 quantiles.

A type of statistical analysis closely related to the estimation of parameters is the testing of hypotheses concerning the values of the parameters. For example, one may wish to test the simple hypothesis that a normal population with a known mean has a variance of σ_1^2 against the alternative hypothesis that the variance is σ_2^2. A number of tests of this nature have been devised using up to four quantiles. As in the case of estimators, one naturally seeks to maximize the efficiency of the test (relative reliability) as defined above. As one might expect, it turns out that for the tests which are concerned with the mean value, the optimum quantiles are those which maximize $\mathrm{Eff}(\hat{\mu})$, and for the tests which are concerned with the variance value, the optimum quantiles are those which maximize $\mathrm{Eff}(\hat{\sigma})$. Consequently, not only have optimum test statistics been devised for these tests, but suboptimum test statistics have also been devised using the suboptimum quantiles for estimating μ and σ. The mathematical and statistical details are given in References 4 and 5. Also included in these reports are quantile estimators based on the degree of correlation between two normal populations.

Thus, it can be seen from the above discussion that, by using quantiles to obtain certain types of statistical information, not only can a significant amount of data compression be achieved, but, equally important, the reduction in uncertainty which accompanies a large sample size is also retained. The investigation into further statistical uses of quantiles is being continued at the Jet Propulsion Laboratory. We shall describe, hereafter, a system which allows the quantiles to be determined onboard a spacecraft. The simplicity of this system, coupled with the large data compression ratio achievable, makes the use of the quantiler appealing.

DATA COMPRESSION RATIOS

If the quantile data are to be a useful form of data compression, a quantiler must be of simple construction. The quantiler design introduced below is one of extreme simplicity; notably, no arithmetic operations are performed. As is discussed in the section entitled "Theoretical Background of the Quantiler," if quantile data are to be useful, the loss that is involved in using quantiles instead of the entire sample must be considered in relation to the degree of data compression that can be achieved. The theory will be discussed from a practical point of view, as it applies to the quantiler design.

A particle count experiment, the outcome of which has a discrete distribution, or empirical distribution or histogram, is used as an example. Before the quantiles are determined, the data are grouped into a familiar histogram. Quantiles of empirical distributions are called sample quantiles. Figure 4-1 shows a typical histogram with four such sample quantiles of orders 0.067, 0.291, 0.709, and 0.933, for 1024 samples.

The sample quantiles are determined by first deciding how many quantiles are to be computed and transmitted. As few as four sample quantiles convey a great deal of useful information. The answer to the statistical questions, "Which quantiles should they be?" and "How good is the information which is obtained from the four quantiles relative to that obtained from the entire sample?" depends on what one wishes to compute.

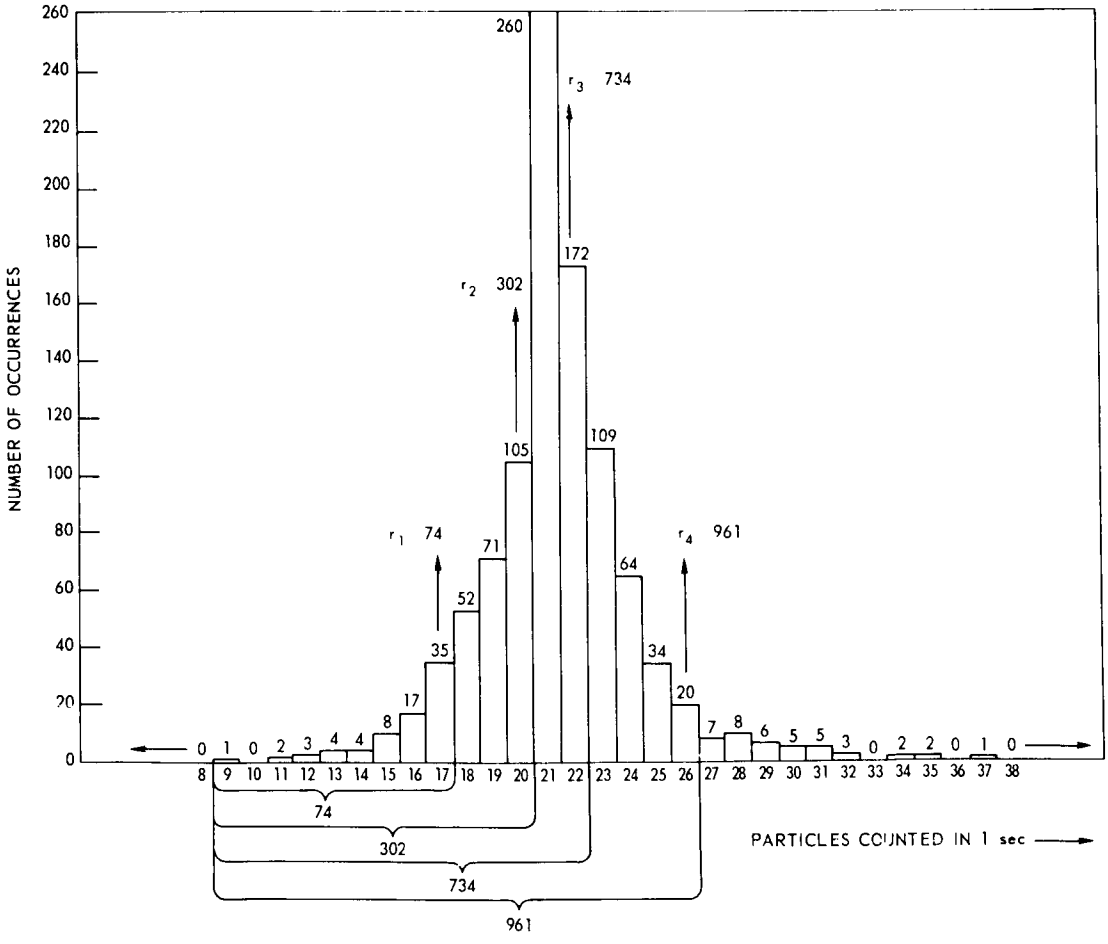


Figure 4-1. A Typical Histogram, 1024 Samples

It should be noted that transmitting quantiles is better than deriving the mean and standard deviation at the test site and then transmitting these moments, even when the greater design simplicity of a quantile system over an arithmetic system is ignored. The knowledge of the quantiles affords additional knowledge of the distribution. For example, the distribution might be bimodal (i.e., have two local maximums) rather than unimodal. Such a condition would be detected using the quantile method but not using the simple mean and standard deviation data. Furthermore, goodness-of-fit tests have been devised for use with quantiles. These are tests which tell an experimenter whether it is reasonable to assume that a given set of experimental data arises from a supposed probability law. This is an important type statistical test in all phases of experimental work. Discriminating between two similar distributions using quantiles extends the applicability of quantiles to data compression still further. Further mathematical details can be found in Reference 2.

The data compression ratios that can be achieved using four quantiles are demonstrated by the following example: The 1024 samples are assumed to be particle counts from 1024 sampling periods. The spread between samples is assumed to be within 128; that is, no more than 127 particles arrive in any one second. A sample size of 1024 is chosen, since it is in the interest of the best estimates of the mean and standard deviation to minimize the variance of the estimates, the variance being inversely proportional to the total number of samples. In fact, the total number of samples should be as large as possible, so long as the incoming data remain stationary. The sample spread should also be large enough to handle a maximum variability in the data rate. Both of these parameters must be limited, however, to hold down the volume of hardware required. As for the sample spread, additional hardware can be more efficiently applied to vary the sampling time, instead of providing extra storage capacity. This self-adaptive feature will be discussed later.

In comparing the transmission of all samples (7×1024 bits) with the transmission of four quantiles (28 bits) per sampling cycle, a compression ratio of 250:1 results. However, consideration of the effect of bit errors for the two types of data gathering and transmission methods leads to the conclusion that the signal-to-noise ratio for the quantile transmission should be increased over that of the raw-data system. Doubling the signal-to-noise ratio, for example, is equivalent to doubling the bit time with constant power; the compression ratio would thus be reduced to 125:1. A further reduction is demanded by the quantile estimation efficiency. For example, with a typical efficiency of 80 percent, the final data compression ratio is found to be in the order of 100:1. By transmitting 1/100 as much data, estimates of the mean and standard deviation of the parent population can be obtained with variances no higher than if the full sample of uncompressed data were used. In effect, this means that 100 percent efficiency is obtained with data compression ratios of 100:1.

PROTOTYPE QUANTILER

One important novelty in the quantiler design is to be found in the method by which the histogram is formed. Other operational functions will be described in reference to the quantiler functional block diagram (Figure 4-2).

Histogram Storage

A quantiler that generates the quantiles for a total of 1024 samples with a sample range of 256 is described in the remainder of this paper. In order to cover all possible sample distributions, prior technology has used 256 registers of $\log_2 1024 = 10$ bits for a total of 2560 bits to store the complete histogram. With the histogram-forming method used by the quantiler, however, only 1280 bits of storage are used. This reduction is accomplished by using a linear storage scheme whereby 1024 spaces (binary zeros) are separated by 256 markers (binary ones) in a serial memory. The markers are used to identify the address of the memory location and the number of spaces between markers represents the number of times that the address appeared as a sample value.

Sampling Rate Controller

The prototype quantiler was built with a particle counting experiment in mind. A source of particles randomly distributed in time was assumed so that the input to the quantiler is merely a counter.

As was stated above, the sample range was 256, which means that an eight-bit counter is used. The input counter must be strobed periodically to sense each data point to be stored. By changing the time between the strobes, varying data rates can be accommodated. The criterion chosen was 16 overflows of the sample counter during the recording of the 1024 sample histogram to double the sampling rate. To accommodate a falling data rate, where perhaps the complete histogram would wind up in only a few lower slots, the highest sample value is observed after the complete histogram has been recorded. If this highest reading is lower than half of the allowable spread, the sampling rate is reduced by one-half before the next histogram is taken. In fact, this automatic control of the sampling rate as a function of the incoming data rate can be considered a secondary data compression feature.

Block Diagram Analysis

Figure 4-2 shows a block diagram of the prototype quantiler. The input random data, derived from a "random pulse generator" (Reference 6), is counted in the data counter and loaded in parallel into the hold register on command from

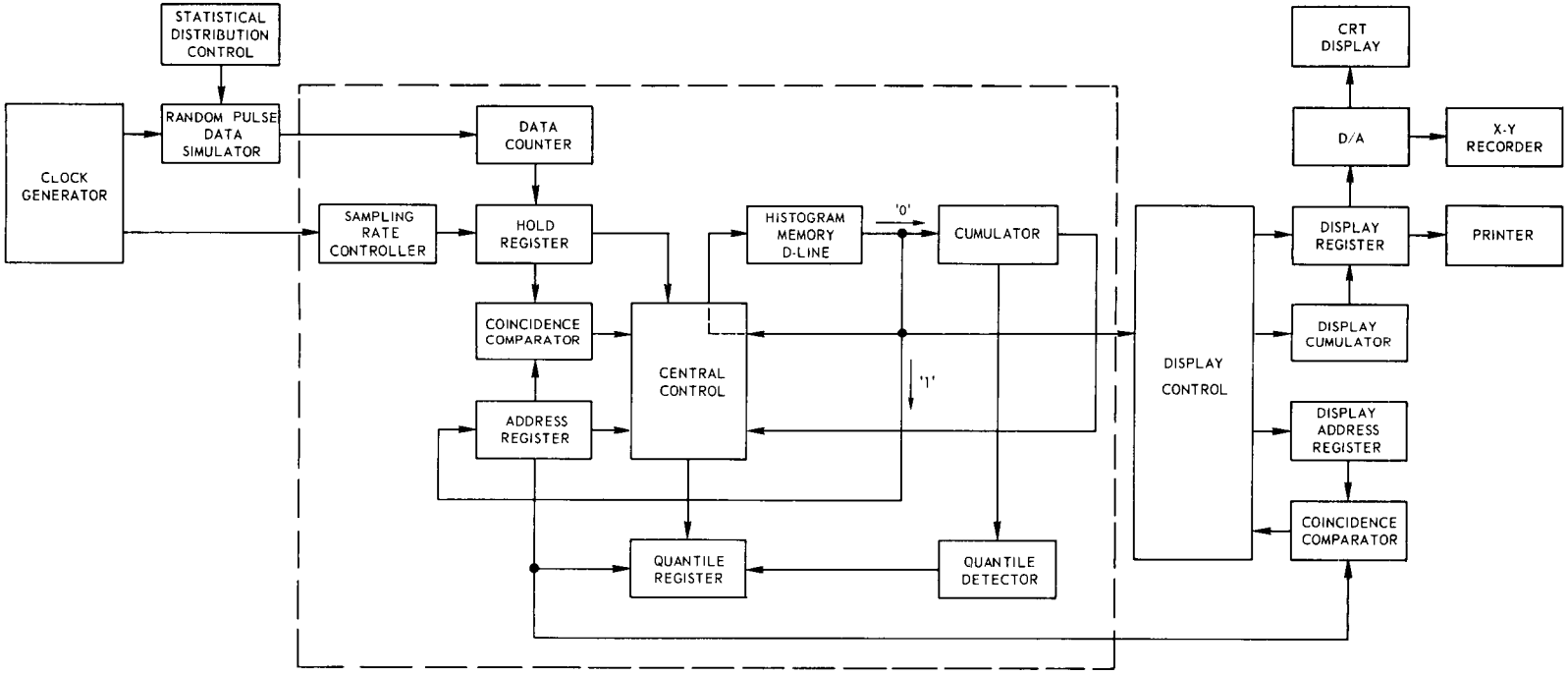


Figure 4-2. Quantiler and Associated Input Output Test Equipment

the sampling rate controller, where it is held until stored in the histogram memory.

The histogram memory is a 1280-bit recirculating delay time. The line is resynchronized between each histogram compilation by storing 256 consecutive ones followed by 1024 zeros. An eight-bit address counter is used to count the ones during synchronization and during data loading to keep track of the address at which the data is to be stored. A cumulator of ten bits counts the 1024 zeros inserted during preload and also counts the samples taken in the histogram store mode. As the delay line recirculates, the address counter counts the ones as they pass. When the address counter reaches the value stored in the data hold register, a coincidence circuit sends a command to the central control to insert a zero just after the binary one that caused coincidence, and then to delay the remaining bits of memory by one pulse. This is done by temporarily adding one bit to the delay line until the last address marker has passed the loading point. Since there are nothing but zeros from this point back to the first address marker, deleting the extra bit at this time will result in losing a zero, thus restoring the line to the required 1024 zeros and 256 ones. This process is repeated serially 1024 times, removing the 1024 zeros from the end of the line and rearranging them between the ones according to the manner in which the data behaved.

After the cumulator counts the 1024^{th} sample, the machine goes into the "quantile compute" mode. In this mode the cumulator counts the zeros in the delay line while the address counter still counts the ones. Since sample quantiles are percentage points of the histogram area, and this area has been constrained to be constant (namely, 1024), the quantiles can be computed simply by counting a fixed number of samples. The four quantiles are thus hard-wired into four different comparator circuits. When the cumulator count equals the value of the desired quantile, a command is given to transfer the value of the address register into the quantile register, where it is then ready for a data buffer for the telemetry.

Display Devices

The prototype quantiler was built with a CRT display of the histogram and with the cumulative function and the four quantiles displayed in octal. Figures 4-3 and 4-4 are tracings of the pictures obtained from a binomial distribution and a bimodal distribution. It can be noted that, even with only four quantiles, the bimodal nature of Figure 4-4 is obvious from the quantiles only.

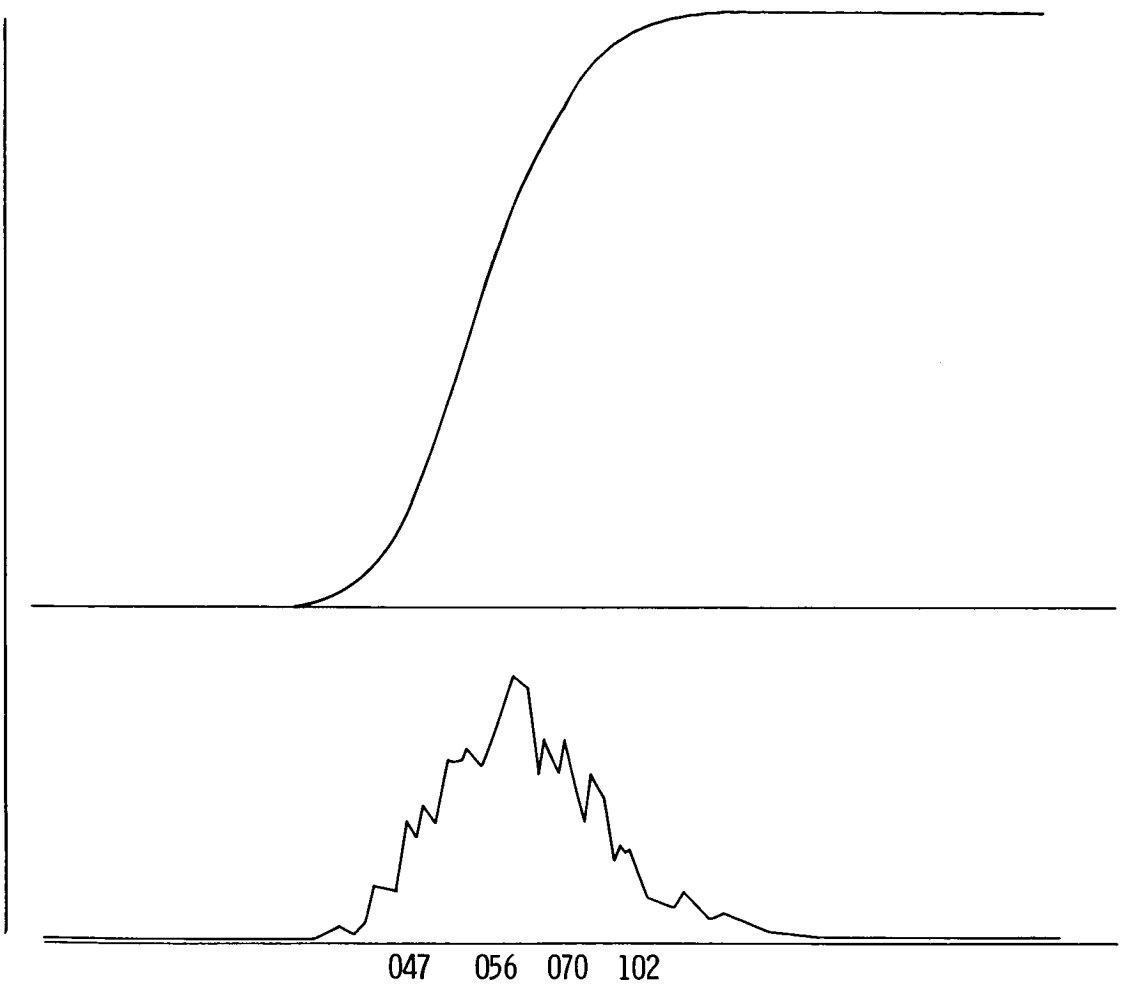


Figure 4-3. Tracing Of Picture Taken Of A Binomial Distribution

Experimental Results

Tests were performed to determine the mean and standard deviation of known distributions by sampling and computing these parameters from the observed quantiles. The estimates of these parameters were calculated by the formulas (taken from the first section)

$$\hat{\mu} = 0.141 \left[Z(0.0668) + Z(0.9332)\right] + 0.359 \left[Z(0.2912) + Z(0.7088)\right] ,$$

and

$$\hat{\sigma} = 0.258 \left[Z(0.9332) - Z(0.0668)\right] + 0.205 \left[Z(0.7088) - Z(0.2912)\right] .$$

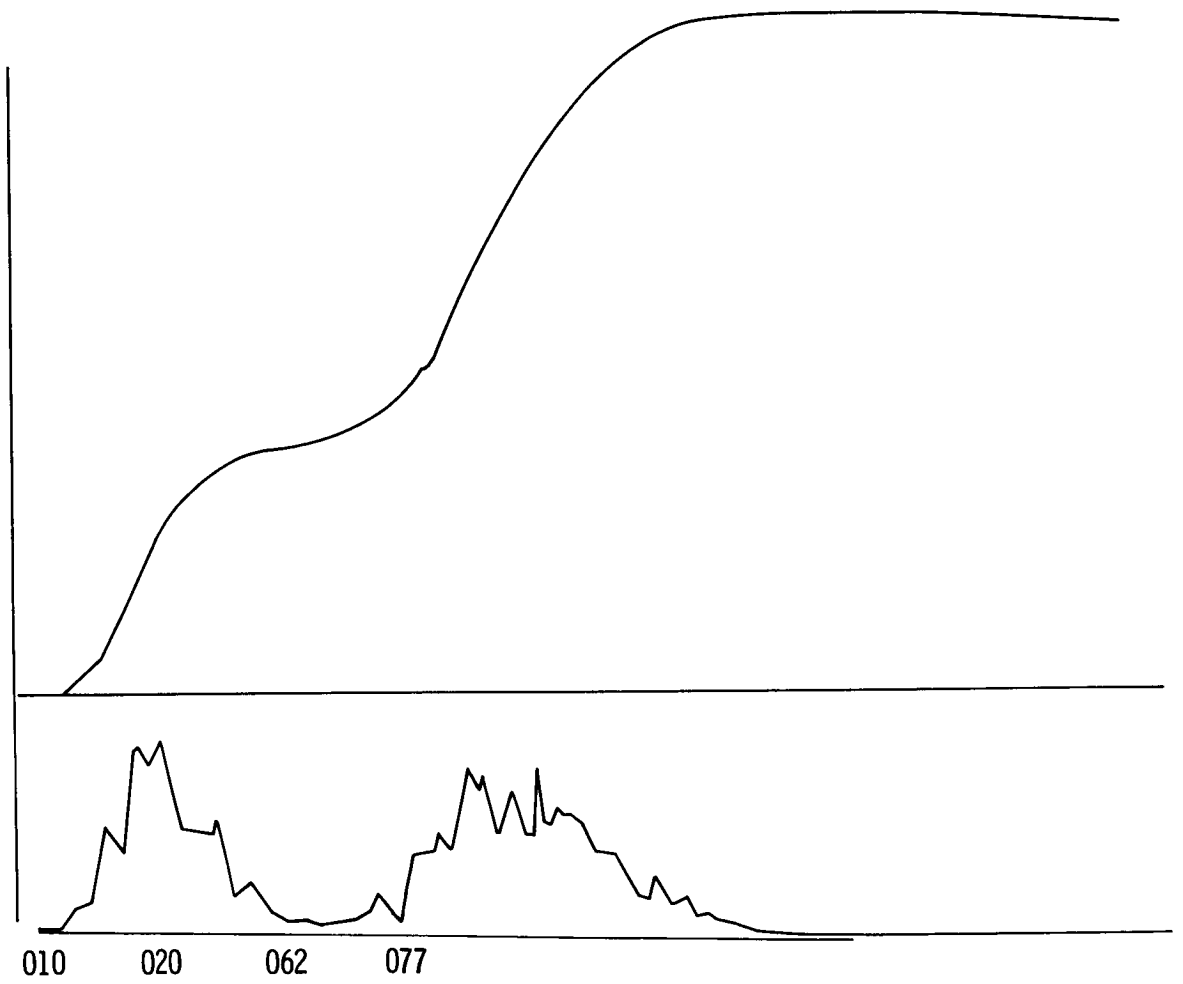


Figure 4-4. Tracing Of Picture Taken Of a Bimodal Distribution

The results of these tests are given in Table 4-1. It is observed that in all cases the agreement between the observed and the theoretical values is excellent. In fact, in the case of the "renewal" data, the variance calculated from the quantiles disclosed an error in the calculation of the theoretical variance.

APPLICATION OF THE QUANTILER

Although the motivation for developing the quantiler was the need for data compression on deep space flights, the relevance of quantile technology to other nonspace telemetry situations became obvious. One use is in industrial quality control. Suppose a product is being manufactured automatically in an inaccessible location. Many quality control techniques use the mean and variance of certain features of the manufactured product to detect departures from the state of

Table 4-1

COMPARISON OF EXPERIMENTAL AND THEORETICAL RESULTS

Known Distribution	Sample Quantiles				$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$
Binomial $\mu = 50$	43	47	52	59	49.9	5.15		
$\sigma = 5.0$	43	47	52	58	49.8	4.89	49.8	4.90
	43	47	52	57	49.6	4.64		
Binomial $\mu = 125$	109	118	130	140	124.1	10.46		
$\sigma = 10.825$	109	118	131	142	124.8	11.18	124.5	10.71
	109	119	130	141	124.6	10.51		
Overlapping Windows - 111	104	116	133	149	125.1	15.10		
$\mu = 125$	103	116	133	149	124.9	15.35	124.99	15.29
$\sigma = 15.31$	104	116	132	151	125.0	15.41		
Overlapping Windows - 101	106	117	132	149	125.4	10.00		
$\mu = 125$	106	116	131	145	125.1	10.51	125.22	10.25
$\sigma = 10.45$	106	117	132	147	125.2	10.25		
Renewal Process - 111	61	67	75	82	71.9	7.06		
$\mu = 71.43$	61	67	75	83	71.3	7.32	71.47	7.23
$\sigma = 7.19$	61	67	75	83	71.3	7.32		

"statistical quality control." A quantiler at the data source would compute the quantiles for transmission to the quality control engineer. From these quantiles, the engineer would receive information not only of the mean and variance but also of other features which may be even more important in quality control. For example, if a bimodal distribution is detected, it could indicate that a machine malfunction occurs part of the time, changing some statistical characteristics of a certain fraction of the output.

Another possible application is the metering of utility usage to determine utility bills. Some utilities charge on a basis not only of average use but of fluctuations about average. If the use per half-hour is recorded in the subscribing facility and quantiles are formed locally once a week and transmitted to the utility, a charge can be made based on mean and variance. The rate commissions probably can be convinced to go along with such schemes.

An application to automatic traffic control also exists. At remote sites throughout a city, the spectrum of vehicle velocities is taken. Quantiles of this

velocity spectrum are transmitted to a central computer, which uses the information to monitor traffic. For example, a low average velocity with a large variance indicates an impending traffic jam.

Other areas for civilian use will be mentioned briefly. Workers in highly radioactive environments can use quantiles to monitor the radiation spectrum. The Weather Bureau can use quantiles to give useable information on quarter-hour temperatures at a given location without having to transmit the temperature every 15 minutes.

Even in space applications there are other types of experiments (other than particle count) to which quantiles apply. Energy spectra of incoming particles are distributions and, as such, can be effectively transmitted by quantiles. The same applies to mass spectrograms in planetary surface life-detection experiments.

In fact, quantiles afford an extraordinary saving whenever any kind of curves, spectra, or distributions have to be transmitted. We believe that the potential applications (especially in a technology where remote sites are linked together by computers) are indeed far-reaching.

REFERENCES

1. Cramer, H., "Mathematical Methods of Statistics," Princeton, N.J.; Princeton Univ. Press, 1946, p. 368.

2. Eisenberger, I., and Posner, E. C., "Systematic Statistics Used for Data Compression of Space Telemetry." American Statistical Assoc, 60(309): 97-113, March 1965; also Tech. Rept. No. 32-510, Jet Propulsion Lab., Calif. Inst. Tech., October 1, 1963.

3. Sarhan, A. E., and Greenberg, B. G., eds., "Contributions to Order Statistics," New York: John Wiley and Sons, Inc., 1962, pp. 272-275.

4. Eisenberger, I., "Tests of Hypotheses and Estimation of the Correlation Coefficient I using Quantiles I," Tech. Rept. No. 32-718, Jet Propulsion Lab., Calif. Inst. Tech., June 1, 1965.

5. Eisenberger, I., "Tests of Hypotheses and Estimation of the Correlation Coefficient II using Quantiles II," Tech. Rept. No. 32-755, Jet Propulsion Lab., Calif. Inst. Tech., September 15, 1965.

6. Anderson, T. O., and E. C. Posner, "Design and Construction of a Random Pulse Generator," Space Programs Summary No. 37-29, Vol. III, Pasadena, Calif.: Jet Propulsion Lab., September 1964, pp. 92-95.

Session II

ADAPTIVE DATA CONTROL AND PROCESSING

Chairmen B. Martin
JPL

R. Stampfl
GSFC

Contributors R. Hollenbaugh
GSFC

J. Sciulli
GSFC

C. Laughlin
GSFC

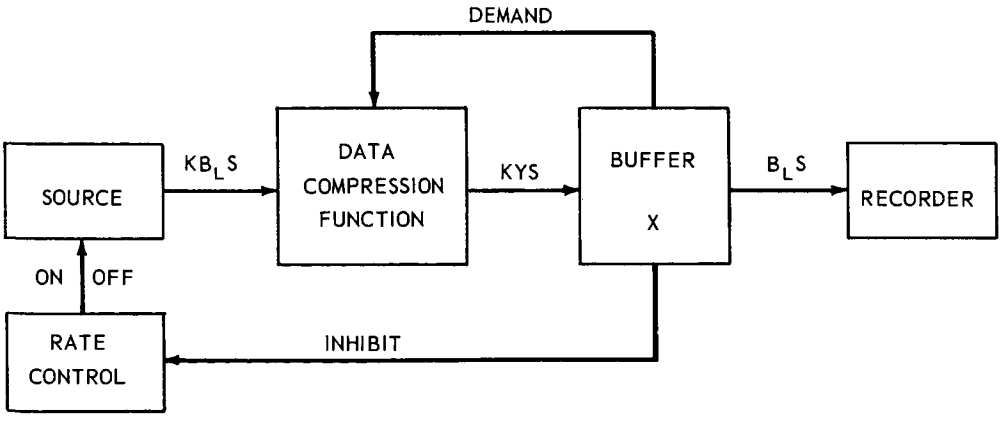
5. A STUDY OF DATA COMPRESSION BUFFERING

R. Hollenbaugh
Goddard Space Flight Center
Greenbelt, Maryland

In a data compression system which removes redundancy from a digitally encoded source, a rate interfacing problem may be encountered. Since the compression operation depends on the data statistics, the output bit rate of the compressor will vary with time. If it is assumed that the source rate does not change and that the total number of bits required to encode the source can be successfully reduced, the average bit rate out of the processing equipment will be reduced. If the equipment receiving the output of the compressor requires a constant bit rate input, an interfacing function will be necessary.

The simplest method of achieving rate compatibility is to place a buffer memory between the data compressor and the readout equipment. Before the readout process is started, the buffer memory is filled. The relationship between the buffer readin rate (determined by the success of the compression operation) and the buffer readout rate (which shall be a constant) determines whether the buffer is being filled or emptied.

Figure 5-1 is a block diagram of a system utilizing a buffer memory for rate interfacing. The two control lines originating from the buffer prevent buffer overflow and buffer emptying. The INHIBIT control sets the source rate to zero when the buffer memory is full. The DEMAND control is necessary to assure that the buffer does not empty before the source is exhausted.



S = TV Scan Rate in Lines per Second.

B_L = Bits Required to Encode One Line of TV Data Before Compression.

Y = Bits required to Encode One Line of TV Data After Compression.

K = Increase in Scan Rate for Data Compression Operation.

X = Buffer Size, in Bits.

Figure 5-1—Block diagram of a system utilizing a buffet memory for rate interfacing

The bit compression ratio $C_R = B_B/B_A$ is the ratio of the number of bits required to encode the source before the compression operation B_B to the number of bits required to encode the source after the compression operation B_A. The INHIBIT control does not affect the C_R that the system can achieve. The DEMAND control, however, can decrease the C_R achieved by the system; exercising this loop increases the number of bits required to encode the source. Ideally, then, it is desirable to have a buffer memory large enough so that it is not emptied before the source is exhausted.

The results described herein show how the C_R that the system can achieve is related to buffer size and the source redundancy that the compression operation can remove.

The controlled-rate source will be a TV camera with 1000 lines and 1000 elements per line. Each element will require six information bits to encode it. The system variables are defined at the bottom of Figure 5-1. The readin rate of the recorder is defined in terms of the system parameters.

In the analysis of the system operation the buffer memory is separated from the system and its operation is analyzed in terms of its input/output parameters. For a given distribution of the redundancy within a picture a piecewise analysis can be performed. With the assumption of a constant readin rate, Equation 1 for T_E, the time required to empty the buffer, and Equation 2 for B_E, the total number of bits passing through the buffer before it is emptied, can be derived:

$$T_E = \sum_{n=0}^{\infty} \frac{(KY)^n X}{SB_L^{(n+1)}} \tag{1}$$

and

$$B_E \cong \frac{X}{1 - \frac{KY}{B_L}} \tag{2}$$

The more it is necessary to use the DEMAND control, the greater is the loss in the compression ratio achieved. If $B_A = B_E$, no loss in C_R is realized since the buffer does not become emptied before the source is exhausted. The worst-case distribution of redundancy is that distribution which causes the buffer readin rate to be at its lowest for the longest time. This situation occurs when all of the nonredundant information is located at the beginning of the picture and the remainder of the picture is completely redundant. Equations (3) show the minimum buffer size necessary to avoid a reduction in the system C_R as a function of U', the number of initial lines of nonredundant data:

$$X \geq 20,265 + 6,735U'$$

and (3)

$$X \geq 579,000 - 579U'.$$

(The equation defining the minimum value of X predominates.)

The above equations assume a minimum C_R of 6/7, a maximum C_R of 2000/7, and K = 10.

From $C_R = B_B/B_A$ one can derive

$$C_R = \frac{1}{\left(\frac{F_a}{C_a} + \frac{F_b}{C_b} + \frac{F_c}{C_c} + \cdots\right)}, \tag{4}$$

where F_a is the fraction of the original picture which can be compressed with a bit compression ratio of C_a, F_b is the fraction of the picture which can be compressed with a bit compression ratio of C_b, and so forth.

For a given distribution of redundancy one can calculate the achieved compression ratio C_R by using Equation 4. Using Equation 4 and assuming the worst case distribution of redundancy and K = 10,

$$C_R = \frac{210,000}{6,979U' - X + 21,000} \tag{5}$$

for $X \geq 7000U'$; also,

$$C_R = \frac{6,000,000}{600,000 + 6,400U' - X} \tag{6}$$

for $X \leq 7000U'$.

The compression ratio that could be achieved without the buffering effect is

$$C_R = \frac{6,000,000}{21,000 + 6,979U'}. \tag{7}$$

For the case where the redundancy is evenly spread throughout the picture

$$C_R = \frac{6,000,000}{600,000 - \frac{XC_{RI}}{10}} \tag{8}$$

by use of Equation 6.

In Equation 8, C_{RI} is the compression ratio achieved by the system if no buffering effect occurs. Again we assume K = 10.

In summary, two specific cases of redundancy distribution have been analyzed. It has been determined that the effect of the buffer size is appreciable only for cases of high compression ratios. The value of K has been found to be a powerful influence in the system operation. The buffer size can be much smaller than the picture size and still perform effectively.

6. RESULTS OF ADAPTIVE PREDICTOR STUDIES. II

J. A. Sciulli
Goddard Space Flight Center
Greenbelt, Maryland

INTRODUCTION

In recent years studies of data compression have warranted the attention of many investigators. Since demands for large amounts of scientific data are increasing, methods for more efficient data transmission must be developed. Usually a communications system is designed so that the information source is sampled at a constant rate determined by the most active data periods. During a large percentage of time the data are relatively quiescent and so redundant samples are transmitted. Data compression by prediction is a promising method of redundancy removal and is therefore the subject of many recent studies. A survey of the literature shows that two philosophies are being proffered for the solution to this problem. The first approach might be called the "state of the art" point of view where efforts have been focused on studying well-known, easily implemented techniques such as the zero-order and first-order predictors. References 1 and 2 are typical examples of this point of view. The philosophy of this approach is that the simpler schemes are within "state of the art" spacecraft instrumentation capability and are certainly easier to study and simulate. Those who have chosen this route generally feel that more sophisticated approaches are too complex to have any application value.

The second school of thought has chosen a more sound theoretical foundation in offering a solution to the data compression problem; the work of Balakrishnan (Reference 3) represents this latter philosophy. It is true that this approach at the present time appears to be difficult to instrument for spacecraft use; nonetheless, it is highly desirable to concentrate on the more sophisticated methods, especially since a high degree of onboard data processing capability (e.g., random access memory and arithmetic capability) will be available in the future.

This report is intended to develop part of the work reported in Reference 3 as well as to complement the work reported in Reference 4. It deals with the description, simulation, and analysis of the results of the application of the conditional expectation predictor to the compression of video data and presents observations on alternate methods and possible applications of the findings. Suggestions for future study are given in the concluding section.

THEORY OF ADAPTIVE PREDICTION SYSTEM

Before describing the prediction mechanism it would be worthwhile to state a few definitions and observations. The word "adaptive" implies modification to meet new conditions. A truly adaptive system is characterized by its ability to (1) monitor its own performance with respect to some performance criterion, (2) learn of new conditions, and (3) adjust its structure to fit the new conditions. In a real communications system no a priori knowledge of the statistical structure of the information source is available. The data compression technique to be described in this report satisfies the definition of adaptivity and also requires no a priori statistical knowledge of the data.

Consider a sequence of discrete samples of the form shown in Figure 6-1. Assume that each sample may take any one of Q discrete values. Suppose a random variable X is defined such that

$$X = (x_{i-M}, x_{i-M+1}, \cdots x_{i-1}). \tag{1}$$

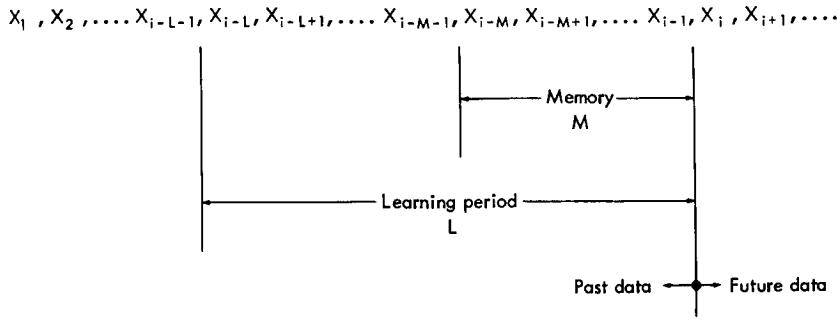


Figure 6-1—Sequence of discrete data samples

The sample space size of the random vector X depends on the choice of the memory size M. Since each sample may assume any one of exactly Q discrete values, the sample space size of X for a memory size M is simply

$$S = Q^M. \tag{2}$$

Suppose, in addition, a second random variable χ is defined such that

$$\chi = x_j \text{ for } 1 \leq j \leq Q \tag{3}$$

and corresponds to the data sample immediately succeeding X. Assume that we have been observing and recording the immediate successors to the random variable X over a number of samples denoted by L, the learning period, and that our operation must determine the optimal prediction for the ith sample. The optimal prediction $\hat{x}_i$ is given by

$$\hat{x}_i = E[\chi / X = (x_{i-M}, x_{i-M+1}, \ldots x_{i-1})]; \tag{4}$$

the notation on the right-hand side of the first equal sign is read as the "conditional expectation of χ given $X = (x_{i-M}, x_{i-M+1}, \cdots x_{i-1})$." In the case of discrete data $\hat{x}_i$ is given simply by

$$\hat{x}_i = \sum_{j=1}^{Q} x_j P[\chi = x_j / X = (x_{i-M}, x_{i-M+1}, \cdots x_{i-1})], \tag{5}$$

where x_j is a possible successor to X and $P[\chi = x_j / X = (x_{i-M}, x_{i-M+1}, \cdots x_{i-1})]$ is the probability that $\chi = x_j$ given $X = (x_{i-M}, x_{i-M+1}, \cdots x_{i-1})$. If the data are assumed to be a long sample from an ergodic process, Equation 5 represents the "best" RMS predictor, since the mean is that point about which the second moment is minimized.

To illustrate by example, assume that k observations of a particular X are made and that at each observation the value of the immediate successor to X is recorded. Suppose a prediction is required for the immediate successor to the $(k+1)$st observation of this particular X. According to Equation 5 the optimal prediction is the mean of the sample of past successors to X and is given by

$$\hat{x}_{k+1} = \sum_{j=1}^{Q} x_j \frac{k_j}{k}, \tag{6}$$

where x_j is a possible successor to X, $1 \leq j \leq Q$, and k_j is the number of times x_j was observed.

Actually one could choose a statistic other than the mean and correspondingly minimize some prediction error criterion other than the mean square error. The mode, for example, could be used as the prediction for the immediate successor to the random variable X. Utilizing the mode as the predictor minimizes the probability of error. In order to implement the mode predictor, histograms representing the distribution of the immediate successor to each particular X in the sample space are constructed. The most frequent successor then becomes the prediction.* One could also choose the median as the prediction; choice of the median minimizes the absolute error. It is interesting to note that if the data were both Gaussian and stationary then the mode, median, and mean would produce identical prediction results.

COMPUTER SIMULATION OF CONDITIONAL EXPECTATION PREDICTOR

The results of this work were obtained from simulations on the IBM 7094 computer using Tiros TV cloud-cover picture data as the information source. Reference 4 contains a good deal of background information on these data, including their origin and subsequent formating for computer simulation. A Tiros TV picture is nominally a 500-scan-line picture with each line composed of 500 TV picture elements. This study has been made on 10 meteorologically significant Tiros TV cloud-cover pictures (Figures 6-12 through 6-21); these pictures are the same pictures as those used for the study reported in Reference 4. Results have been obtained with each TV element quantized to 4 and 6 bits.

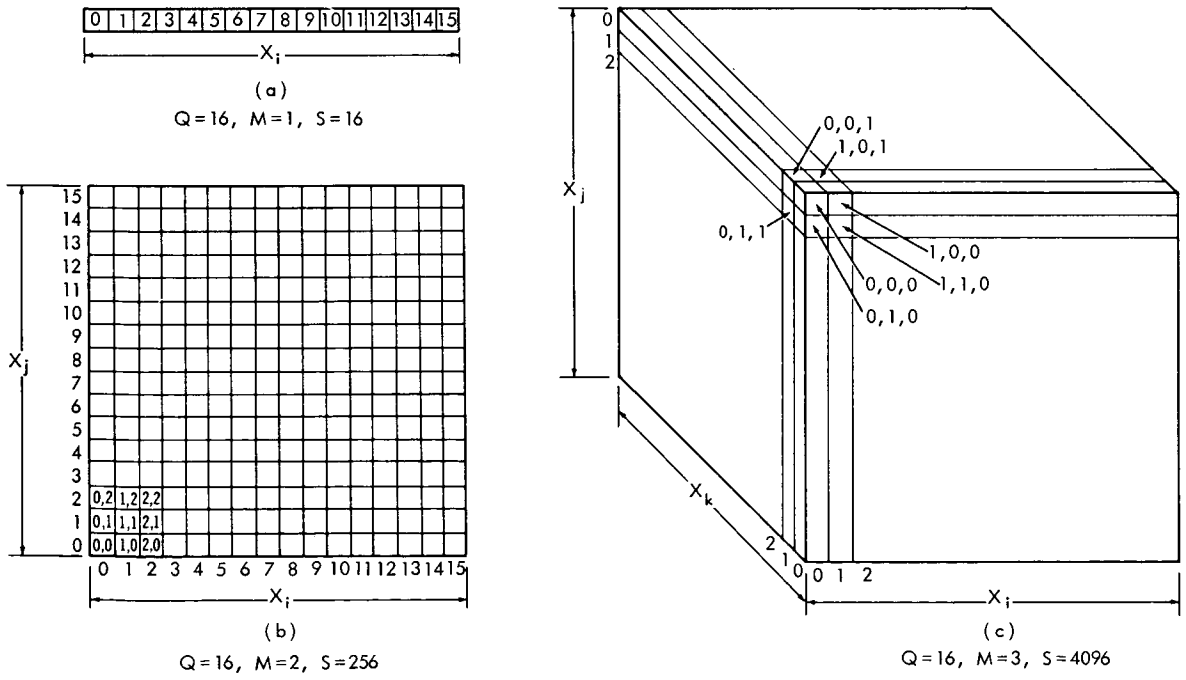


Figure 6-2–Memory cell geometries for Q = 16 and M = 1, 2, and 3

*This technique was implemented by Davisson of Princeton during his participation in the 1965 Goddard Summer Workshop (Reference 5). His results in some cases were somewhat better than those using the mean as the prediction.

Assume that the video data is to be scanned one element at a time from left to right and top to bottom, beginning with the top leftmost TV element. The choice of the parameter M (memory size) determines the number of M-dimensional cubes (called M-cubes in this paper) which are required to store the statistical structure of the data. For example, if the data are quantized to 16 levels Q and a memory size M of 2 is chosen then there must be exactly Q^M or $(16)^2 = 256$ 2-cubes required. Figure 6-2 shows memory cell geometries for $Q = 16$ and $M = 1, 2$, and 3.* The process begins by scanning the data one element at a time and observing the random variable X. At each observation of X the prediction for its immediate successor is computed from the statistics stored in the M-cube associated with the particular X under observation. The prediction error is given by

$$E_p = |x_a - x_p|, \tag{7}$$

where x_a is the actual value and x_p is the predicted value. If $E_p \leq T$ where T is a preset allowable error threshold, the element is predictable and need not be transmitted. If, however, $E_p > T$ this particular element is not predictable and must be transmitted in unmodified form.

The data compression system at the transmitter end must provide the receiver with the data necessary to reconstruct the original message within the allowable error threshold T. To accomplish this the predictor at the transmitter end must operate on exactly the same data which it will send to the receiver for reconstruction. Therefore, if an element is predictable ($E_p \leq T$) it need not be transmitted, but the predicted value is treated as though it were the actual value and is also used to update the statistics stored in the M-cube defined by the X under observation. If, however, an element is not predictable ($E_p > T$) the actual value is used to update the statistics. This is called "closed-loop" operation. The prediction mechanism could be evaluated in the "open-loop" mode. In open-loop operation predicted values do not replace actual values, thus the predictor operates on raw data only. The studies described in this report, however, were done in the closed-loop mode.

The example given previously described the formulation of the sample mean in terms of Equation 6. In the computer simulation it is not necessary to keep track of the relative frequency terms k_j/k because each M-cube defined by X can be composed of two storage locations, a sum location and a counter location. At each observation of X, the sum location corresponding to this X is updated by adding to the existing sum either the actual or predicted value of the successor to X depending on whether the element is predictable. At the same time the corresponding counter location is incremented by one count for each observation of X. The k_j terms of Equation 6 are implicitly contained in the sum at all times. Therefore the prediction computation need be performed only when a prediction is required and is easily obtained by dividing the sum by the counter.

Because the learning period includes only a finite amount of past data, a prediction for the successor to a particular value of the random variable X could frequently be indeterminate because of a complete lack of past information; this is especially true at the beginning of the learning process. One solution might be to determine a prediction from the statistics contained in the M-cubes neighboring the particular M-cube defined by the X under observation. This approach, however, does not solve the problem at the beginning and in the very early stages of the learning period. The obvious solution then is to make some initial assumption for the successor to each of the values which the random variable X can assume before the learning process begins. If it turns out that the initial assumption was a poor one it will affect the efficiency of the prediction mechanism less and less significantly as more and more of the data are observed. This scheme was utilized in the simulation of the conditional expectation predictor and the choice of the prelearning assumption was made based on the results of the zero-order hold predictor (Reference 4).

*For memory sizes larger than 3, the number of storage locations required becomes unwieldy for practical computer simulations.

Experiments with the zero-order hold predictor showed that very frequently an element intensity was within ±1 or ±2 quantum levels of its predecessor. Thus, if the random variable X associated with the conditional expectation predictor is a 2-dimensional random vector

$$X = (x_r, x_{r+1}), \tag{8}$$

the prelearning assumption for the successor to the $(r+1)$st element is the $(r+1)$st element. Similarly, if $X = (x_{r-1}, x_r, x_{r+1})$, the prelearning assumption would again be x_{r+1}.

Quite often a suitable prediction cannot be derived from the statistics contained in the particular cube defined by the X under observation. When this occurs it is possible to utilize the neighborhood statistics as the source of a secondary prediction. For example, (Figure 6-3) suppose the random variable under observation is

$$X = (i, j) \qquad 1 \leq i \leq Q, \ 1 \leq j \leq Q,$$

where i and j are values of element intensity specifying the coordinates of a specific 2-cube in the memory array. Suppose that the conditional expectation calculated from the statistics contained in (i, j) is inadequate; that is, $E_p > T$. As soon as it is determined that the prediction error E_p exceeds the threshold T, a secondary prediction is provided by computing the mean of the statistics contained in the 2-cubes in the neighborhood of cube (i, j). The boundaries of the neighborhood are governed by the allowable prediction error threshold T so as to accommodate the fidelity criterion. For example, if T is ±1 quantum level and a suitable prediction cannot be made from cube (i, j), the cubes which are not more than ±1 quantum level away from (i, j) are those from which the secondary prediction is determined. The concept of providing a secondary prediction if the primary prediction fails is in itself attractive, but this attractiveness is somewhat dulled when one considers that the use of alternate prediction modes in the same compression mechanism complicates the coding problem since the receiver must determine the source of each prediction.

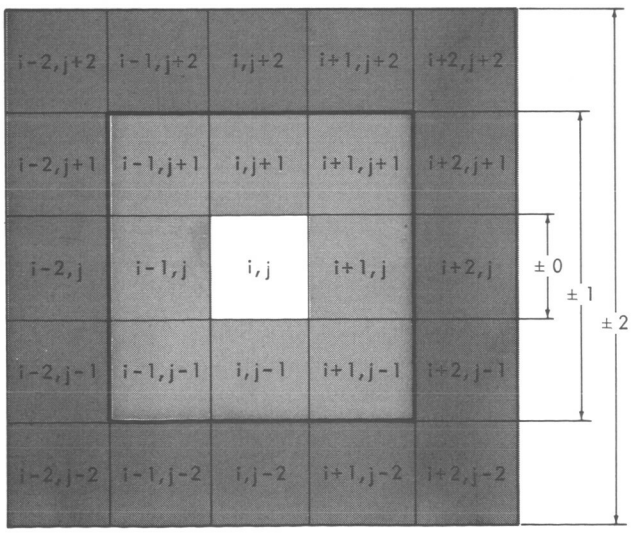


Figure 6-3—Two-dimensional memory cube and its neighboring cubes

The discussion thus far assumes that the TV data are observed serially one element at a time, scanning from left to right. There is some advantage, however, in observing the data not only from left to right along a TV line but also from line to line so as to take advantage of the vertical correlation in the TV data. Figure 6-4 depicts the geometry of the TV data. If one wishes to operate the prediction mechanism only on data scanned serially from left to right, the random variable X would take the form of an ordered pair of adjacent elements on the same line; e.g., typically, $X = [x_{i,j}, x_{i,j+1}]$. If, however, one wishes to take advantage of line-to-line correlation, X might consist of an ordered pair of TV elements of the form $X = (x_{i,j}, x_{i+1,j+1})$ where the element to be predicted is $x_{i,j+1}$. This scheme might be termed an elementary, two-dimensional predictor.

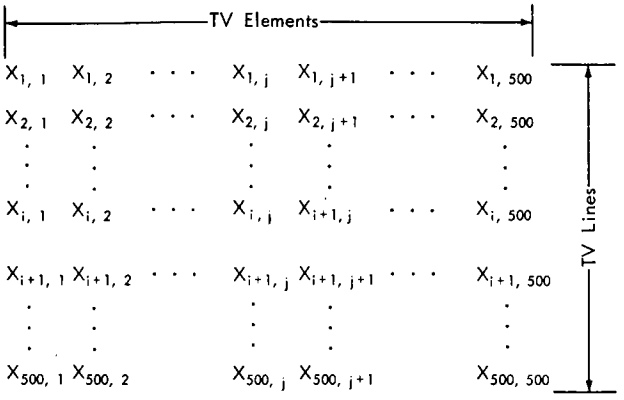


Figure 6-4—Geometry of TV data

So far not too much has been said about the learning operation. Actually, the learning operation of the conditional expectation predictor (Method II of Reference 3) is not so explicit as that of the linear predictor (Method I of Reference 3). In Method I the learning period is composed of about 20 data samples preceding the elements to be predicted. The function of this learning period is to develop an optimal operator based on these 20 previous points. In Method II, however, the function of the learning period is to determine the optimal operation to predict the successor to the present observation of the random variable X. This is the basic difference between Method I and Method II. Method I determines an optimal operator based on a few points preceding the elements to be predicted, while Method II determines the optimal operation based on previous observations of the successor to the particular X under observation. Also, for Method I either a linear or a nonlinear operation is explicitly chosen. For example, Method I as it is described in Reference 4 is very obviously linear. Method II, however, does not distinguish between linear and nonlinear operations. The conditional expectation predictor simply proceeds to the optimum operation without restriction to either linear or nonlinear operation.

Since the learning period of the linear predictor is used to determine an operator over a fairly small number of previous data samples and the learning period of the conditional expectation predictor is used to observe occupancies of a relatively large number of M-cubes, it seems reasonable that the second method should require a much larger learning period than that required by the first method. Results from computer simulations included in the discussion of results appear to support this argument. It is important to note that in Method I the optimal operator is found over a learning period just preceding the sequence of elements to be predicted and a new learning process does not begin until the mean square prediction error exceeds a preset threshold. In Method II, however, the learning process is more continuous in nature and prediction and learning take place almost simultaneously.

Method I as described in Reference 4 utilizes two thresholds. The first is the threshold T, which is the allowable error between true and predicted values of a data sample. The second threshold is associated with the mean square prediction error which is calculated periodically to determine the prediction ability of the present operator. When the mean square prediction error exceeds this threshold, the prediction mechanism is signaled to restart its learning operation. Thus far in the description of Method II only one threshold has been mentioned. This is the threshold T which corresponds exactly to the first threshold of Method I. Method II in its present configuration does not employ a second threshold equivalent to that of Method I.

In the first simulation of the conditional expectation predictor the learning-period length was chosen based on parametric trials with the element compression ratio serving as the figure of merit. Figure 6-5 shows these results with the data quantized to 6 bits per TV element, the memory M = 1 TV element, and an allowable error threshold T of ±2 quantum levels, where cumulative element compression ratio is the 4800-line (10-TV-picture) average compression ratio. This is not an ideal way to handle the learning operation. It might be worthwhile to monitor the mean square prediction error and introduce a second threshold as in Method I. The problem with this, as in the present configuration, is that the start of a new learning period causes a large instantaneous drop in the amount of statistical data available with which to make predictions.

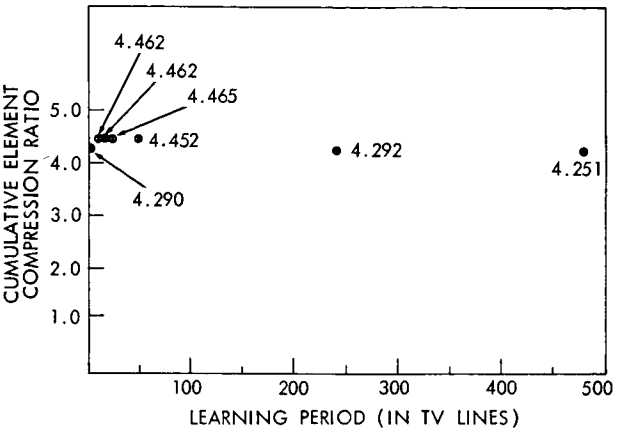


Figure 6-5–Cumulative element compression ratio versus learning period (in TV lines) with Q = 64 (6 bits/element), M = 1, T = ±2 quantum levels

A solution free from this problem is to allow the statistical structure to decay slowly to some effective N element average. Each M-cube of the memory array is composed of a summer and a counter. Suppose the counter is allowed to build up freely to N observations and future observations are handled as follows: Let

σ_N = Sum contained in the sum location after N observations

and

ρ_{N+1} = (N + 1)st sample.

Then at the (N + 1)st observation σ_N is replaced by

$$\sigma_{N+1} = (\sigma_N + \rho_{N+1}) \left(\frac{N}{N+1}\right) \tag{9}$$

Furthermore,

$$\sigma_{N+2} = (\sigma_{N+1} + \rho_{N+2}) \left(\frac{N}{N+1}\right) = (\sigma_N + \rho_{N+1}) \left(\frac{N}{N+1}\right)^2 + \rho_{N+2} \left(\frac{N}{N+1}\right),$$

$$\sigma_{N+3} = (\sigma_{N+2} + \rho_{N+3}) \left(\frac{N}{N+1}\right),$$

and so on. Thus the most recent observation is weighted most significantly; the second most recent observation, the second most significantly; and so forth.

DISCUSSION OF RESULTS

Figures 6-12 to 6-21 are copies of the Tiros TV cloud-cover pictures used in this study. These 10 pictures are the same as those used in the study reported in Reference 4. The background of the original analog data, the construction of the unmodified digital pictures, and the description of the display of these same pictures after processing with the prediction mechanism are also contained in Reference 4. The pictures which appear in this report probably will have lost some of the linearity of the gray scale because of the reproduction process but their overall quality should not be degraded because of the large number of gray scales present.

The complete data compression system embraces two problems, the prediction problem and the coding problem. Although they are not independent, it is possible to think of them as two distinctly separate problems. In order to separate them one needs to impose the constraint on the prediction mechanism that it at least does not hamper the coding mechanism in reasonably representing the data. With this consideration in mind the results obtained so far can be presented in two parts. The first section will deal with the characteristics of the prediction mechanism with element compression ratio as the standard of comparison. The second section presents some possible approaches to the coding problem with bit compression ratio as the standard of comparison.

Results of Simulations of Prediction Mechanism

Since it was assumed that the prediction problem and the coding problem were separate, the objective of the simulation of the prediction technique was to maximize the element compression ratio. Element compression ratio is defined as the ratio of the total number of TV elements in the original unmodified picture to the total number of unmodified TV elements which must be transmitted after the picture is processed by the prediction mechanism. Element compression ratio then is simply a measure of the "predictability" of the data and certainly does not include coding considerations. The objective of the initial work was to simulate the technique and evaluate the results with the element compression ratio serving as the figure of merit.

Learning Period Considerations

Table 6-1 shows element compression ratios for the basic prediction scheme with the data quantized to 6 bits per TV element, $M = 1$, $T = \pm 2$ quantum levels, and learning periods varying from 2 TV lines to 480 TV lines in one picture. There certainly are no significant gains in compression ratio for any of the learning periods used. However, as a first choice one might pick a learning period length of 256 samples (approximately 1/2 TV line) for this case. The reasoning for this is quite simple. Consider the general case with a memory size M, with the data quantized to Q quantum levels. As described earlier in the report, prediction depends on the conditional

Table 6-1–Various Learning Periods with Q = 64 (6 Bits/Element); M = 1 and T = ±2

Figure Number	Element Compression Ratio for Learning Period L of —						
	480 TV lines	240 TV lines	48 TV lines	24 TV lines	16 TV lines	10 TV lines	2 TV lines
12	5.478	5.740	5.970	5.978	5.935	5.964	5.601
13	4.814	4.820	5.023	5.024	5.006	4.967	4.725
14	4.322	4.432	4.725	4.720	4.709	4.699	4.479
15	4.873	4.935	5.063	5.082	5.094	5.036	4.878
16	3.908	3.962	4.053	4.053	4.042	4.062	3.897
17	3.688	3.710	3.822	3.814	3.810	3.793	3.660
18	4.206	4.205	4.329	4.335	4.362	4.400	4.300
19	2.376	2.385	2.450	2.460	2.452	2.447	2.381
20	4.256	4.237	4.461	4.533	4.527	4.559	4.336
21	9.550	9.527	10.103	10.205	10.381	10.453	10.025
Cumulative Element Compression Ratio	4.251	4.292	4.452	4.465	4.462	4.462	4.290

expectation of the successor to a random variable X whose sample space size depends on M. In particular, the sample space size $S = Q^M$. Thus, if $M = 1$, $Q = 16$, and the objective is to predict X_p when X_{p-1} is known, there are exactly $(16)^2$ or 256 possibilities for the set (X_{p-1}, X_p). Therefore, if all cases were equiprobable one would have to allow the learning period to cover 256 samples to be sure that each case was observed at least once. Thus for the general case of Q and M one might choose as the minimum learning period length $L = Q^{(M+1)}$. This certainly does not represent the optimum learning period length but it does provide a quideline as to the minimum learning period length. One might govern the upper bound of the learning period size by investigating the changes in the structure of the statistics as more and more samples are observed. In any case it is advantageous to keep the learning period size as small as possible, since the data are suspected to be highly nonstationary.

The results of successive experiments designed to test the performance of the conditional expectation on both 6- and 4-bit data are contained in Tables 6-2 and 6-3. Figures 6-6 and 6-7 depict these results as bar plots. A few conclusions can be drawn from these results:

(1) There is essentially no difference between the results for $M = 1$ and $M = 2$ without the statistical-neighborhood and two-dimensional prediction modes.

(2) A slight improvement in compression ratio was achieved when the learning period L was reduced from 480 to 16 TV lines.

(3) Significant improvements in compression ratio were achieved with the addition of the neighborhood and two-dimensional predictors.

Table 6-2–Element Compression Ratios with Q = 64 (6 Bits/Element) and T = ±2

Figure Number	Element Compression Ratio for —				
	M = 1; L = 480 TV lines	M = 1; L = 16 TV lines	M = 1; L = 16 TV lines with SNP[1]	M = 2; L = 16 TV lines with SNP[1]	M = 2; L = 16 TV lines with both SNP[1] and EAP[2]
12	5.478	5.935	7.420	8.712	11.425
13	4.814	5.006	5.982	7.120	9.734
14	4.322	4.709	5.686	6.703	8.134
15	4.873	5.094	6.249	7.217	8.444
16	3.908	4.042	4.826	5.521	6.413
17	3.688	3.810	4.424	4.994	5.913
18	4.206	4.362	5.295	6.543	6.863
19	2.376	2.452	2.825	3.316	3.297
20	4.256	4.527	5.469	6.874	7.875
21	9.550	10.381	12.623	15.941	17.550
Cumulative Element Compression Ratio	4.251	4.462	5.330	6.301	7.196

[1] Statistical neighborhood predictor.
[2] Element area predictor.

Table 6-3–Element Compression Ratios with Q = 16 (4 Bits/Element) and T = ±1

Figure Number	Element Compression Ratio for —				
	M = 1; L = 480 TV lines	M = 1; L = 16 TV lines	M = 1; L = 16 TV lines with SNP[1]	M = 2; L = 16 TV lines with SNP[1]	M = 2; L = 16 TV lines with both SNP[1] and EAP[2]
12	7.443	7.846	9.697	9.762	14.267
13	6.355	6.725	8.182	7.995	12.705
14	6.092	6.410	7.777	7.566	10.390
15	6.256	6.721	8.015	7.786	10.356
16	4.904	5.258	6.331	6.523	8.146
17	4.611	4.896	5.575	5.753	7.379
18	5.457	5.588	6.870	7.051	7.824
19	2.930	3.055	3.593	3.750	3.890
20	5.418	6.334	7.789	7.923	9.690
21	13.442	13.449	17.127	15.550	19.954
Cumulative Element Compression Ratio	5.494	5.835	7.009	7.071	8.787

[1] Statistical neighborhood predictor.
[2] Element area predictor.

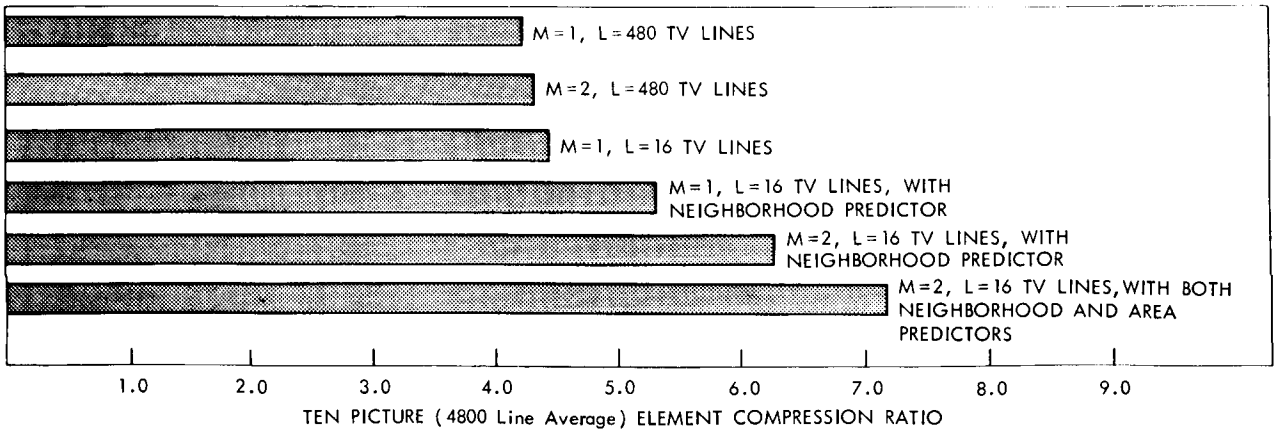


Figure 6-6–Performance bar plot for conditional expectation predictor with Q = 64 (6 bits/element) and T = ±2 quantum levels

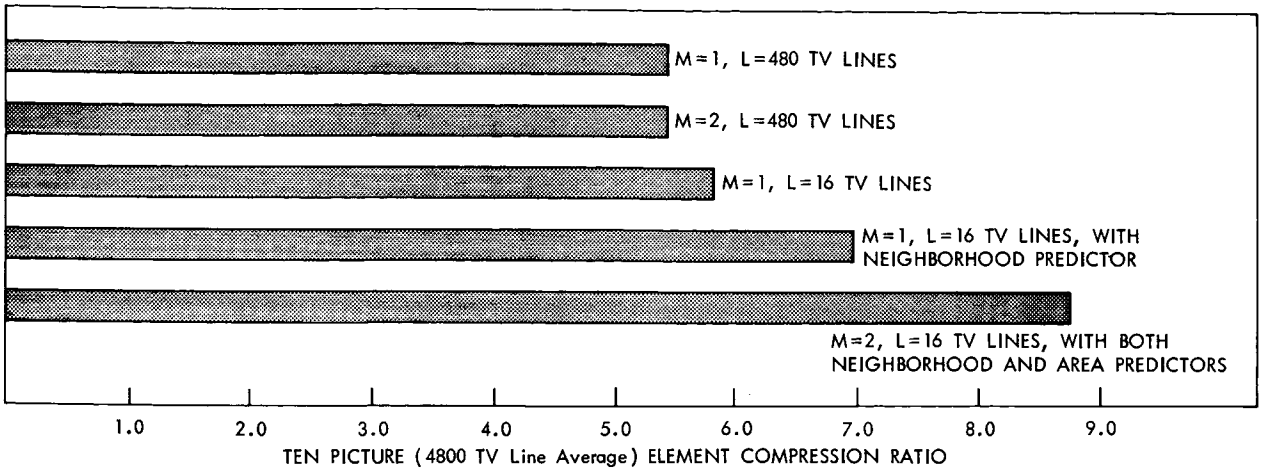


Figure 6-7—Performance bar plot for conditional expectation predictor with Q = 16 (4 bits/element) and T = ±1 quantum level

Comparison with Other Techniques

Figures 6-8 and 6-9 summarize in bar-plot form the relative performance of:

(1) The zero-order hold predictor.

(2) The linear predictor of Reference 4 (Method I - Reference 3).

(3) The conditional expectation predictor (Method II - Reference 3).

The linear predictor (Method I) produces a 10-picture cumulative element compression ratio of about 3:1 for 6 bits per element and T = ±2. The zero-order hold predictor provides a compression ratio of about 4.2:1 for 6 bits per element and T = ±2 and one of about 5:1 for 4 bits per element and T = ±1. The conditional expectation predictor in its most elementary form (without neighborhood and two-dimensional predictors) performs slightly better than does the zero-order hold. The conditional expectation method along with the neighborhood and two-dimensional predictors shows a significant gain with a ratio of more than 7:1 for 6 bits per element and T = ±2 and a ratio of nearly 9:1 for 4 bits per element and T = ±1. One might reason that the zero-order hold does very well with respect to the other two methods cited when the relative complexity of the schemes are considered. The only explanation as to why the zero-order hold predictor does this well is that the information source is very highly nonstationary. Certainly the zero-order hold results would be much less impressive if the information source were nonstationary. Note, however, that the conditional expectation predictor does about 140 percent better than the linear predictor and about 70 percent better than the zero-order hold. One reason that the conditional expectation predictor does so much better than the linear predictor is that the former is not restrictive with respect to linear or nonlinear operations and therefore is able to predict well despite the nonstationary character of the data.

It was mentioned earlier that the incorporation of the statistical neighborhood predictor as an alternate prediction mode contributes to the coding costs since the receiver must determine which statistics were used to make the prediction. One way to overcome this problem would be to constrain the transmitter always to make predictions from the neighborhood statistics. Preliminary results with the technique have shown that the mechanism is not able to predict nearly so well when only neighborhood predictions are permitted.

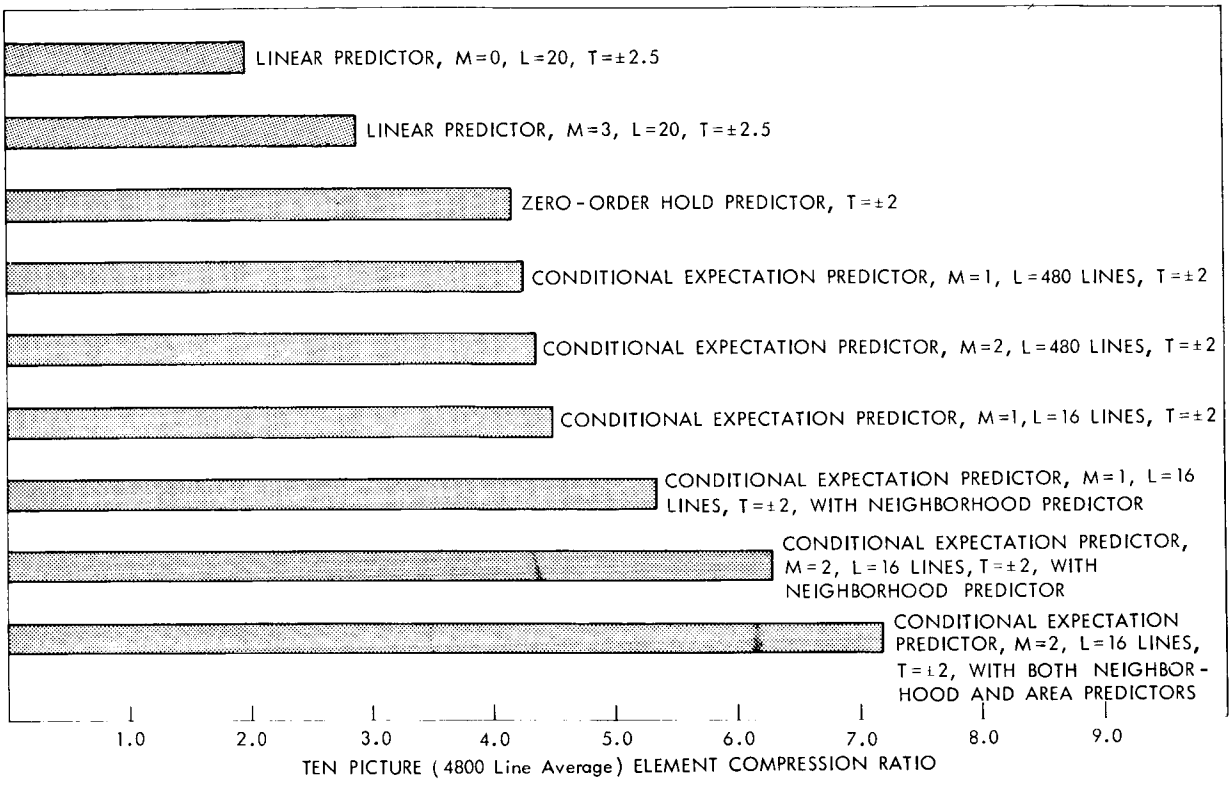


Figure 6-8—Performance of conditional expectation predictor relative to linear predictor of reference 4 and zero-order hold predictor with Q = 6 bits per element.

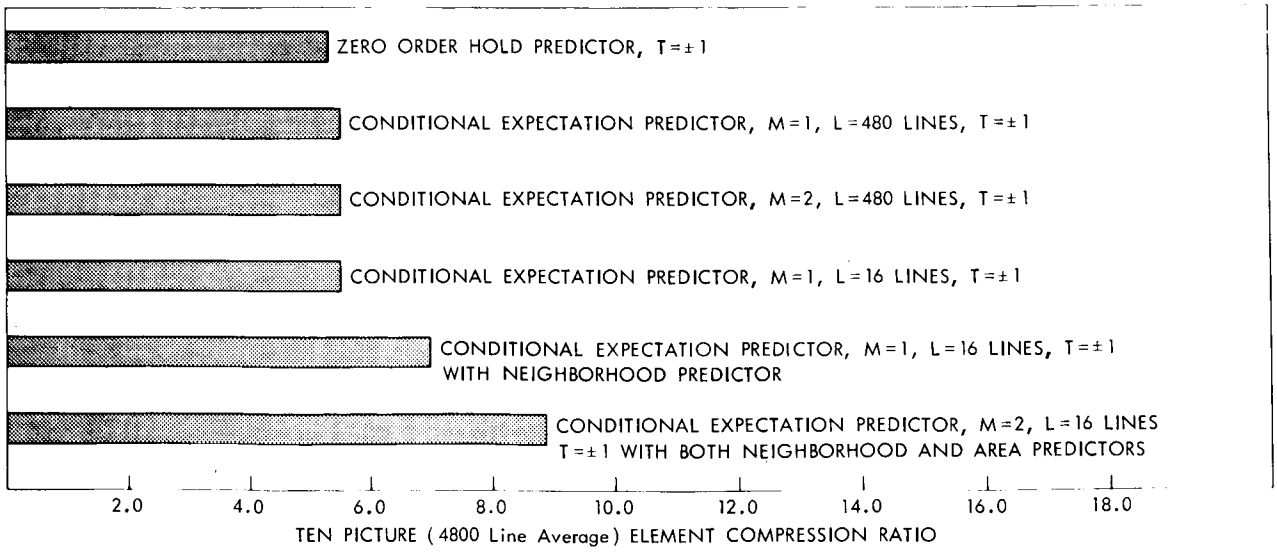


Figure 6-9—Performance of conditional expectation predictor compared with performance of zero-order hold predictor with T = ±1 and Q = 4 bits per element

An explanation for the basis of the choice of the allowable prediction error T seems to be necessary at this point. Obviously the selection of T is very important to the performance of the prediction mechanism. Since the information source here is video data representing cloud-cover pictures the effect of the choice of T can be easily observed when the data processed by the prediction mechanism is displayed. The problem here is that a judgment of the quality of a compressed picture must be made subjectively by eye; thus the only solution is to try different thresholds until the maximum threshold which allows retention of minimum acceptable picture quality is determined. The choices of $T = \pm 2$ quantum levels for the 6-bit case and $T = \pm 1$ quantum level for the 4-bit case were made after experimenting with a number of thresholds. Figure 6-10 is a picture showing the effects of too large a value of T with the data quantized to 64 levels and $T = \pm 4$.

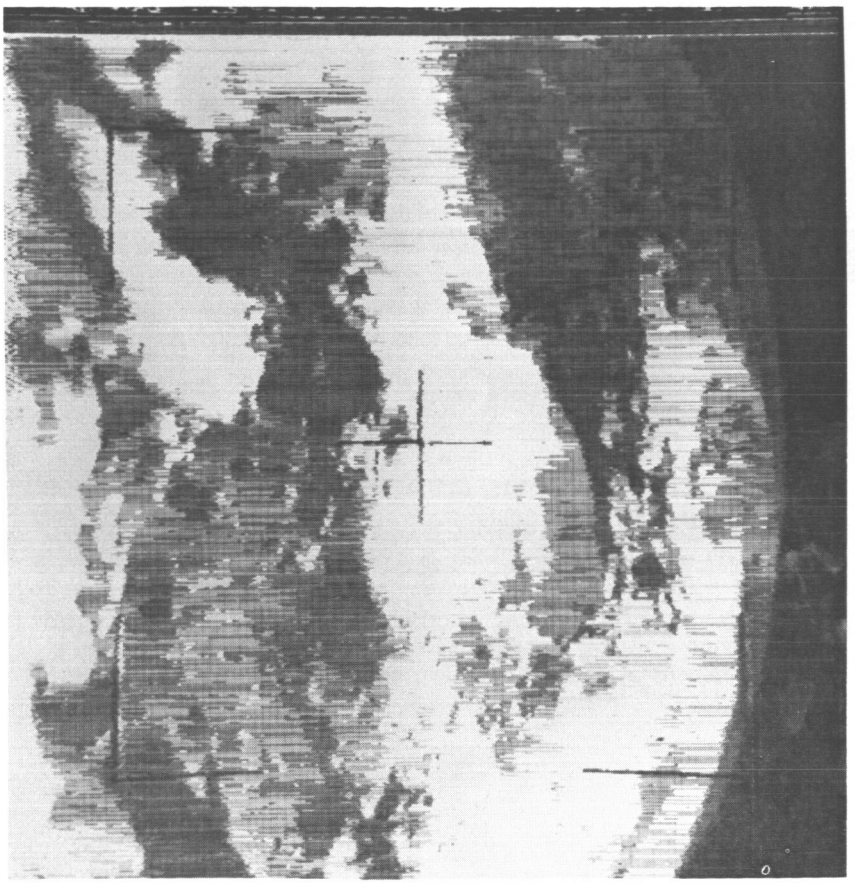


Figure 6-10—Effect of choosing too large a value of T (allowable prediction error). In this case Q = 64 (6 bits/element) and $T = \pm 4$ quantum levels.

Coding Considerations

The prediction problem, while not completely defined, has certainly been investigated more thoroughly than has the coding problem. The most important question is, "After prediction what does the transmitter send to the receiver?" This report will not deal explicitly with the coding problem, but will offer a few observations about it. Actually, the problem of coding for a data compression system is not an easy one, and very little work has been done in this area.

The problem with most standard coding schemes is that they require knowledge of the statistics of the data. The prediction philosophy clearly states that no a priori knowledge of the statistics is necessary. It therefore seems reasonable that the coding philosophy should not be constrained by this requirement either.*

In order to evaluate any hypothesis adequately it is helpful to have some standard of comparison which is optimum in some sense. Suppose that P is the probability of making an accurate prediction and also that each of Q levels is equally likely when accurate prediction is not possible. If it is also assumed that the ability to predict is sample-to-sample independent, then theory explains that in the noise-free case a bit compression ratio (including coding costs) of

$$\frac{\log_2 Q}{P \log_2\left(\frac{1}{P}\right) + (1-P)\log_2\left(\frac{Q}{1-P}\right)} \qquad (10)$$

can be approached with optimum coding. Figure 6-11 is a family of curves of bit compression ratio C_B against element compression ratio C_E with $\log_2 Q = 4$ and 6. The probability of predicting accurately P is related to the element compression ratio C_E by

$$P = 1 - \frac{1}{C_E} \quad . \qquad (11)$$

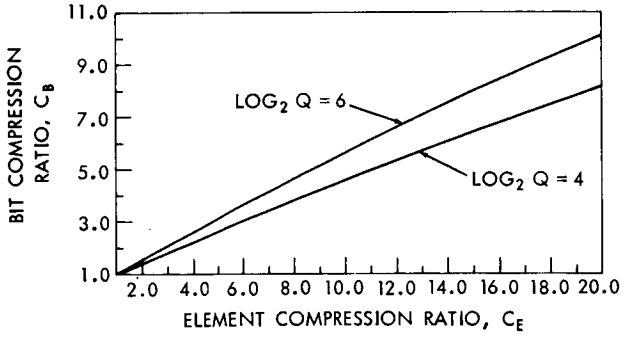


Figure 6-11—Bit compression ratio C_B versus element compression ratio C_E for $\log_2 Q = 4, 6$

Table 6-4—Element Compression Ratios and Corresponding Bit Compression Ratios for Conditional Expectation Predictors

Figure Number	Q = 64 levels; T = ±2; M = 2; L = 16 TV lines		Q = 16 levels; T = ±1; M = 2; L = 16 TV lines	
	Element Compression Ratio	Bit Compression Ratio	Element Compression Ratio	Bit Compression Ratio
12	11.425	6.285	14.267	6.189
13	9.734	5.481	12.705	5.607
14	8.134	4.705	10.390	4.757
15	8.444	4.851	10.356	4.729
16	6.413	3.846	8.146	3.888
17	5.913	3.593	7.379	3.705
18	6.863	4.064	7.824	3.762
19	3.297	2.218	3.890	2.160
20	7.875	4.576	9.690	4.473
21	17.550	9.127	19.954	8.219
Cumulative Compression Ratio	7.196	4.238	8.787	4.136

*Reference 4 shows some examples of coding the compressed data with variations of run-length coding. These results are interesting and similar simulations might be made with the conditional expectation predictor.

Table 6-4 provides examples of resultant bit compression ratios for each of the 10 pictures with Q = 64 and T = ±2 and with Q = 16 and T = ±1. It must be made clear that these results are by no means quotations of bit compression ratios one could obtain in practice for the following two reasons:

(1) The results assume optimum coding which would probably not be attainable in practice.

(2) These results apply to the noiseless channel and do not account for necessary error-correction coding.

These data are presented solely to provide guidelines to those who demand results which are in line with practical arguments.

Comments on the TV Pictures

Each of Figures 6-12 to 6-21 contains in the following order:

(1) A photograph of the original analog picture.

(2) A photograph of the original digital picture constructed from the analog data.

(3) Two photographs of the digital data redisplayed after processing by the conditional expectation predictor.

Reference 4 contains a good deal of information on the history and specific characteristics of many of these pictures, as well as a description of the techniques used to display them.

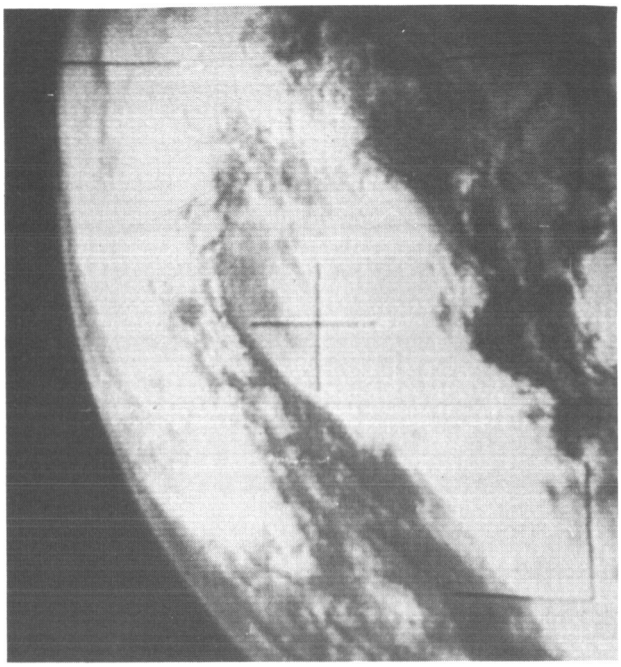
The author will not attempt to give a detailed meteorological analysis for each picture but will rather provide a general comparison of the pictures processed by the conditional expectation predictor with the unmodified pictures as well as with those processed by other compression techniques. It is impossible for the untrained eye to pass judgment as to the retention of meteorological fidelity of the compressed pictures.* The only alternative for the layman is to compare the compressed pictures subjectively with the originals and to estimate the loss of apparent picture quality.

When the pictures processed by the conditional expectation predictor are compared with the digital originals, the loss of picture quality is obvious but not objectionable. Contouring or "streakiness" in highly detailed regions seems to be the most popular complaint. This contouring is caused by the ability of the prediction mechanism to predict long sequences of elements at the same level successively. This effect becomes more pronounced as T is increased. A technique which might partially solve this problem is the use of a weighted prediction error criterion where the prediction errors are accumulated until a present threshold has been exceeded.**

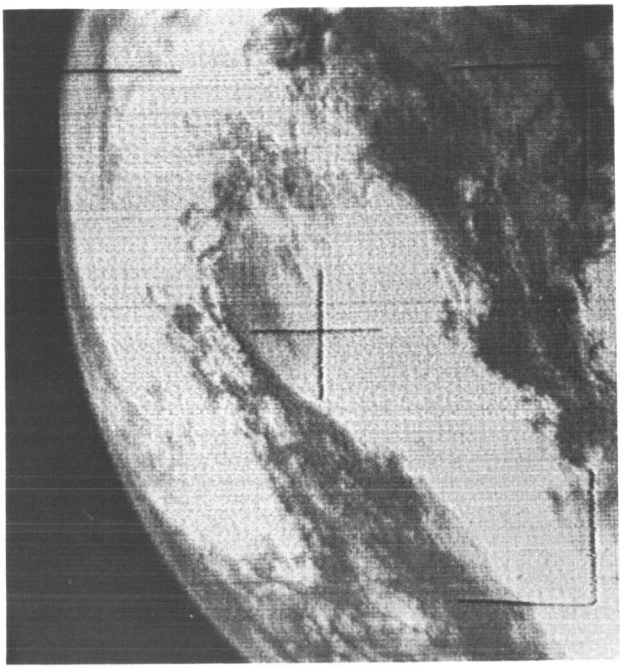
The pictures quantized to 4 bits per element with T = ±1 (Figures 6-12(c) to 6-21(c)) exhibit a higher degree of picture quality degradation than do the pictures quantized to 6 bits per element with T = ±2 (Figures 6-12(d) to 6-21(d)). The reason for this is that a threshold of ±1 quantum level at 4 bits per element is a larger percentage error than a threshold of ±2 quantum levels at 6 bits per element. Both the 6-bit and the 4-bit pictures are displayed with 16 shades of gray. The 6-bit compressed pictures with T = ±2 are acceptable although the 4-bit compressed pictures with T = ±1 seem to be at the threshold of acceptability. Perhaps the best compromise would be

*The term "compressed" picture does not imply that the picture geometry is made smaller or more compact in any way. It is simply true that the amount of data required to transmit a "compressed" picture over a communications link is less than the amount of data required to send the original picture.

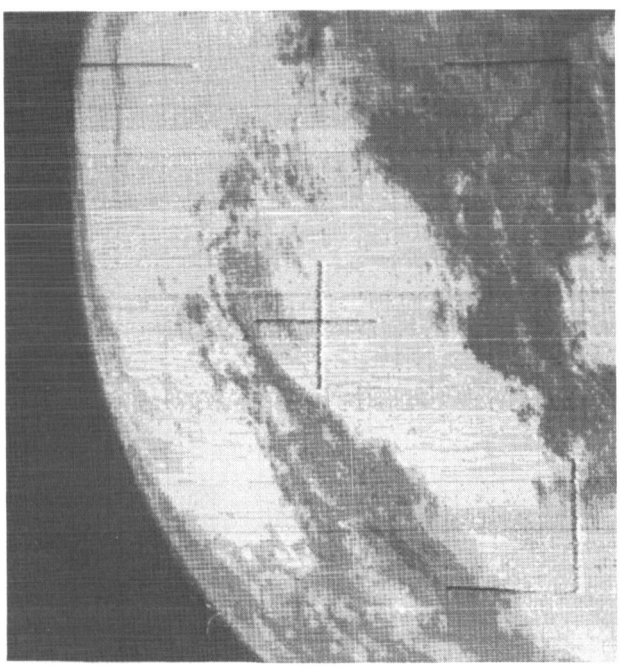
**This scheme was implemented by Davisson of Princeton and described in the report of his work in the 1965 Goddard Summer Workshop (Reference 5).



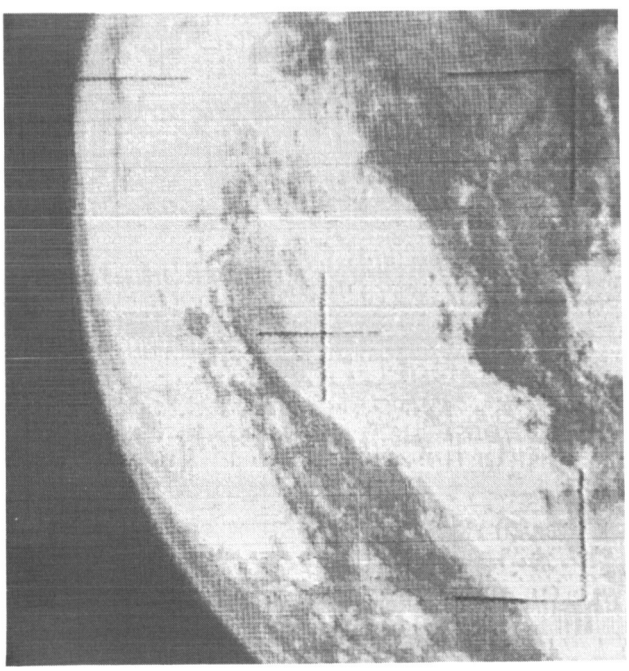
(a) Analog original.



(b) Digital original.

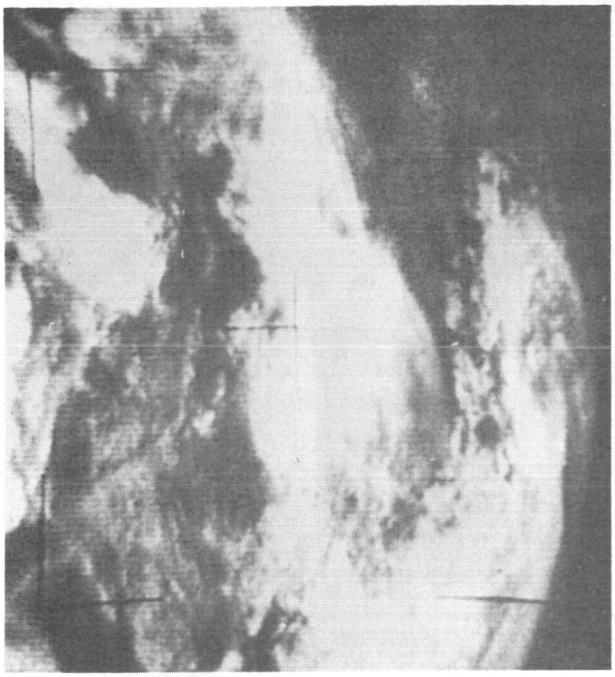


(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 14.267; bit compression ratio, 6.189.

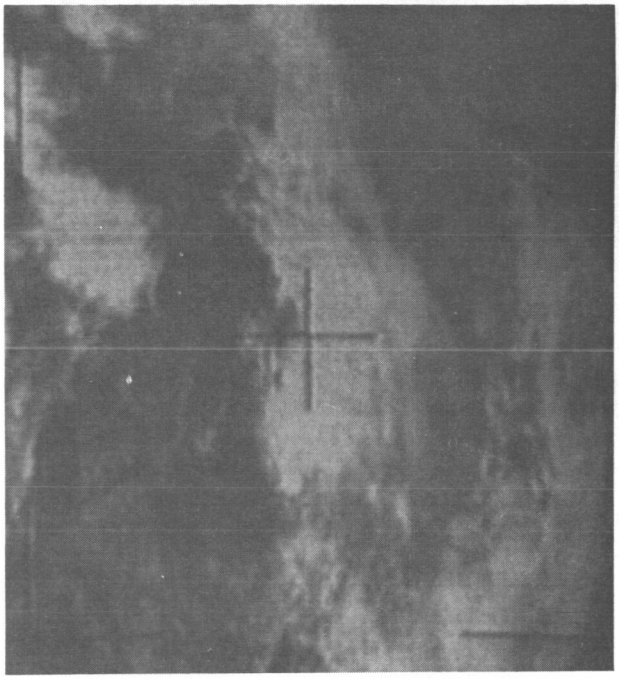


(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 11.425; bit compression ratio, 6.285.

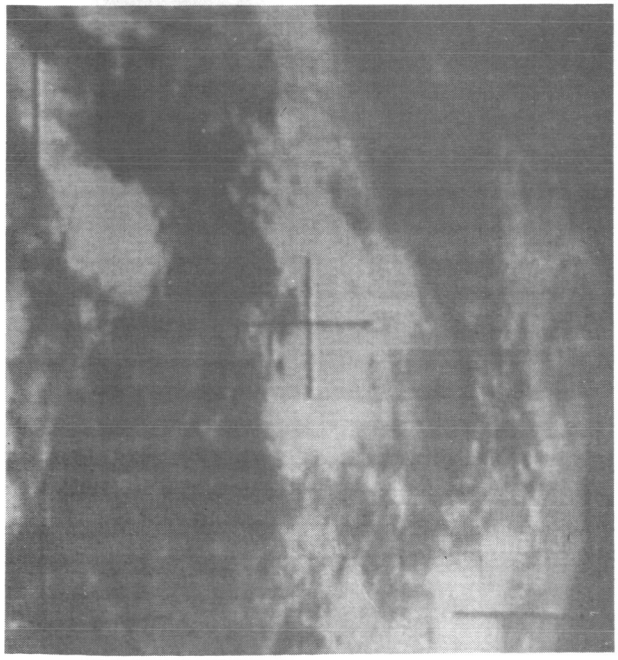
Figure 6-12—Pictures from Tiros III, orbit 4, frame 2, camera 2; direct transmission from satellite; principal point, 43.6N, 95.5W; subsatellite point, 41.0N, 89.2W



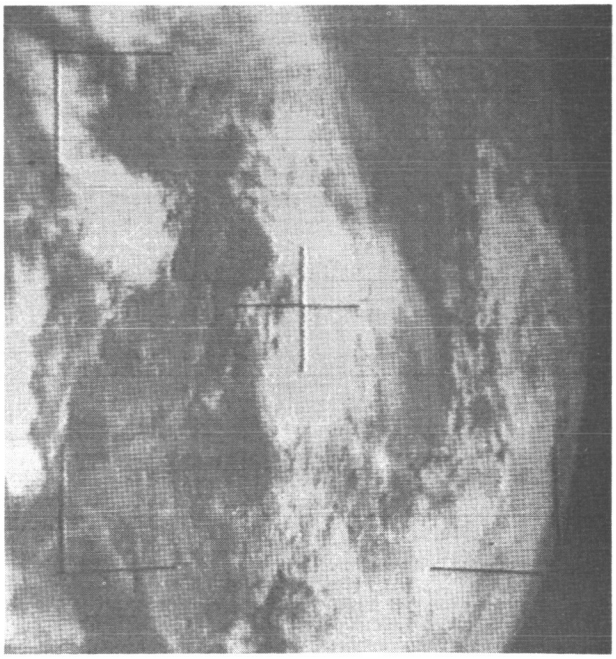
(a) Analog original.



(b) Digital original.



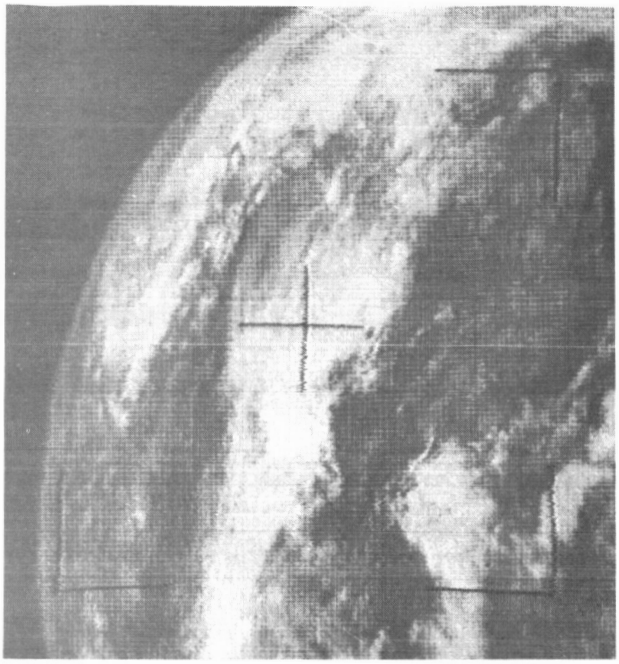
(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 12.705; bit compression ratio, 5.607.



(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 9.734; bit compression ratio, 5.481.

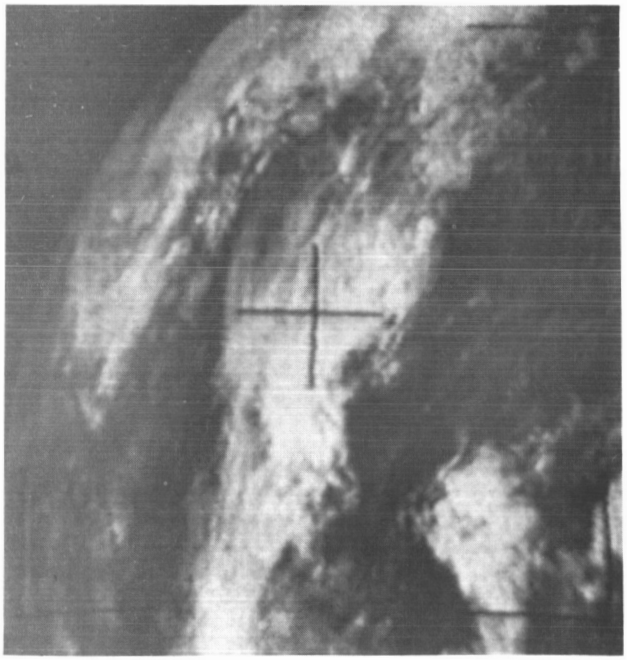
Figure 6-13—Pictures from Tiros III, orbit 4, frame 3, camera 2; direct transmission from satellite; principal point, 43.4N, 95.0W; subsatellite point, 40.8N, 88.8W

(a) Analog original.



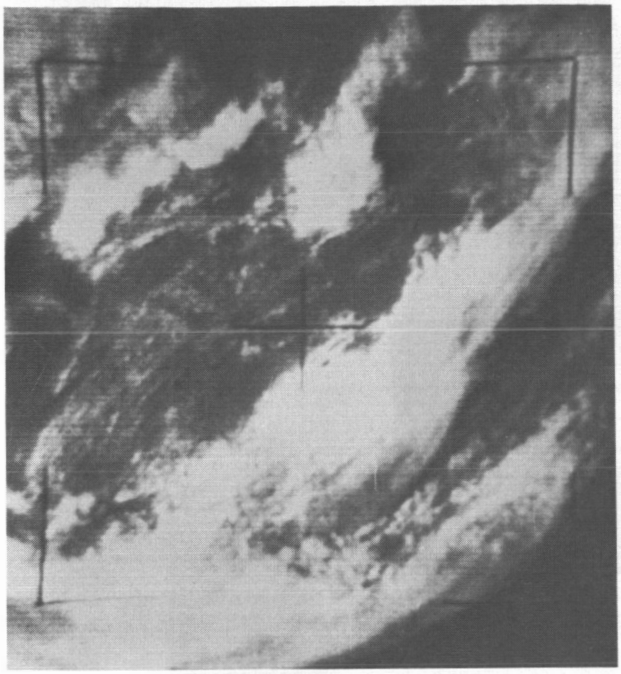
(b) Digital original.

(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 10.390; bit compression ratio, 4.757.

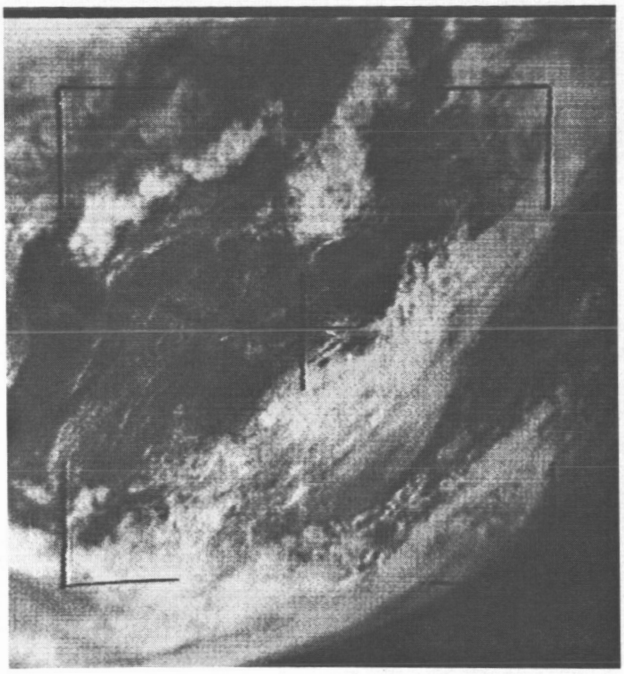


(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 8.134; bit compression ratio, 4.705.

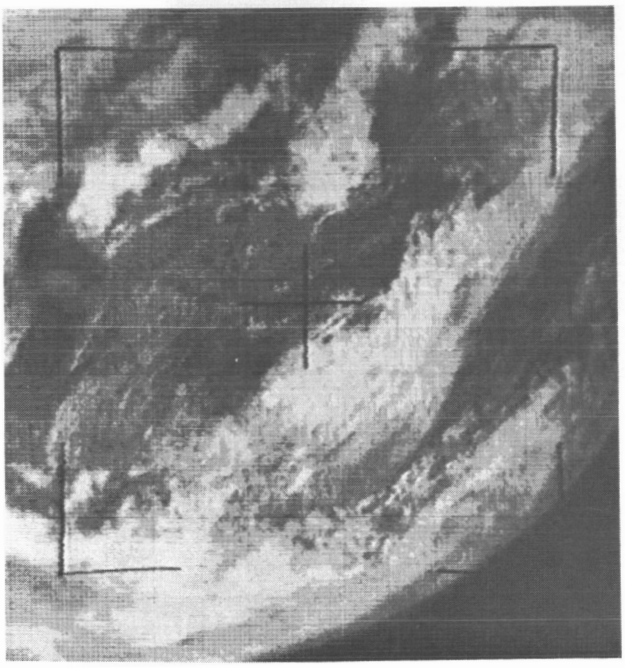
Figure 6-14—Pictures from Tiros III, orbit 4, frame 4, camera 2; direct transmission from satellite; principal point, 43.0N, 94.0W; subsatellite point, 40.5N, 88.1W



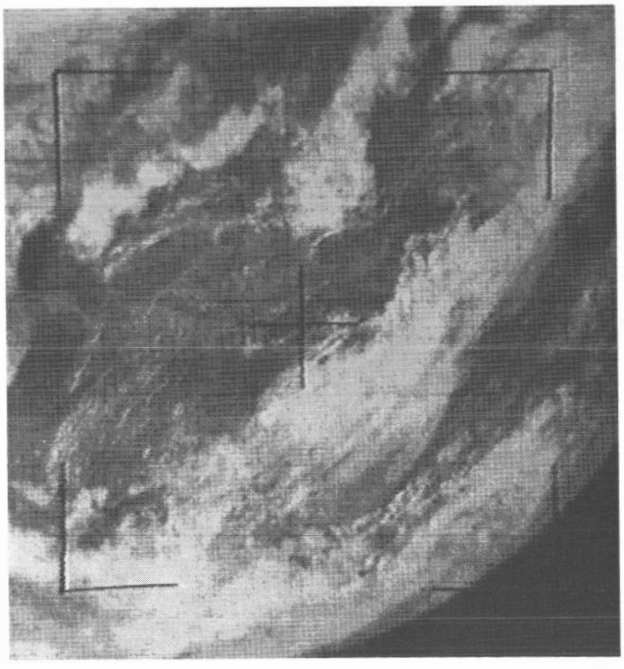
(a) Analog original.



(b) Digital original.

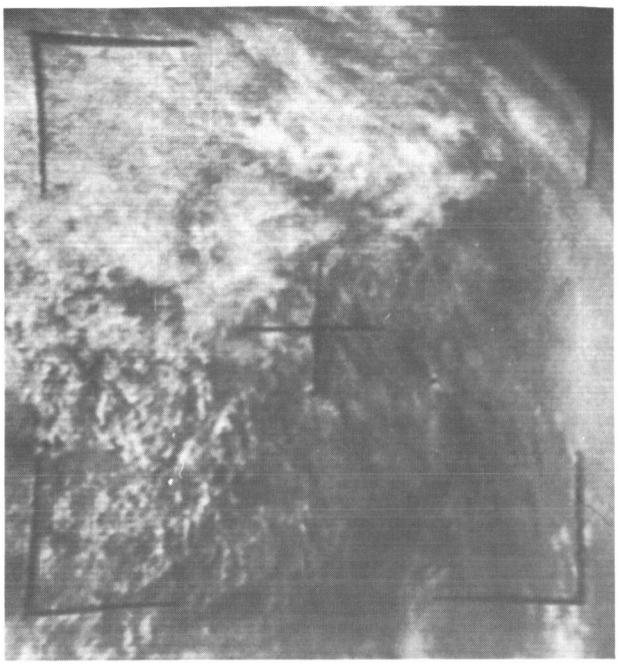


(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 10.356; bit compression ratio, 4.729.

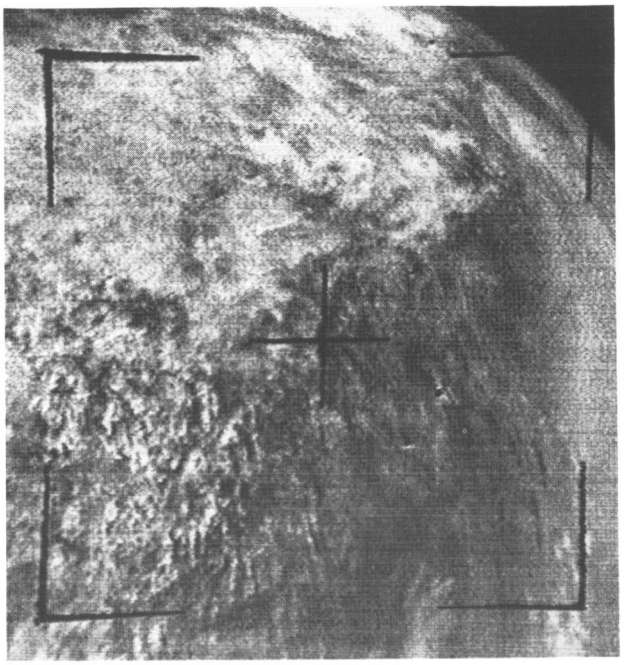


(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 8.444; bit compression ratio, 4.851.

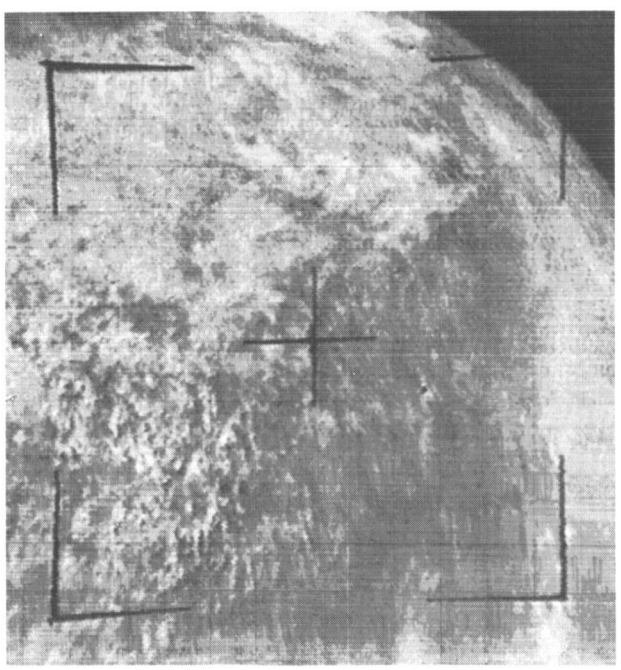
Figure 6-15—Pictures from Tiros III, orbit 4, frame 5, camera 2; direct transmission from satellite; principal point, 42.6N, 93.0W; subsatellite point, 40.1N, 87.3W



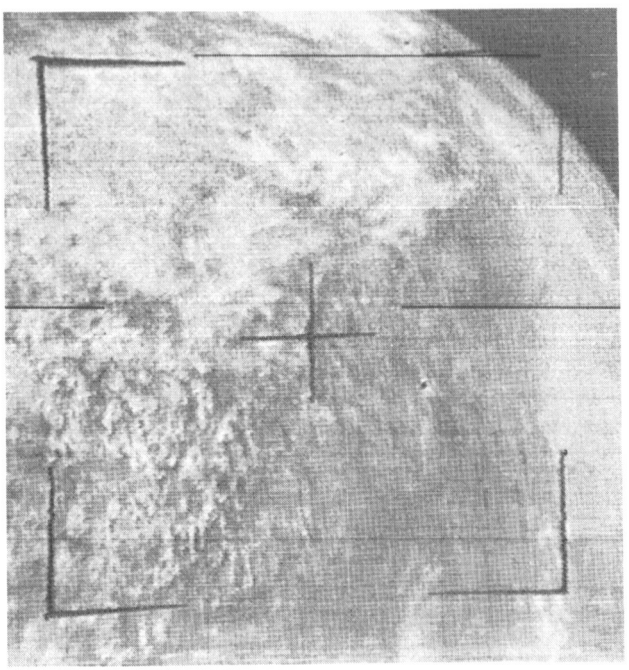
(a) Analog original.



(b) Digital original.

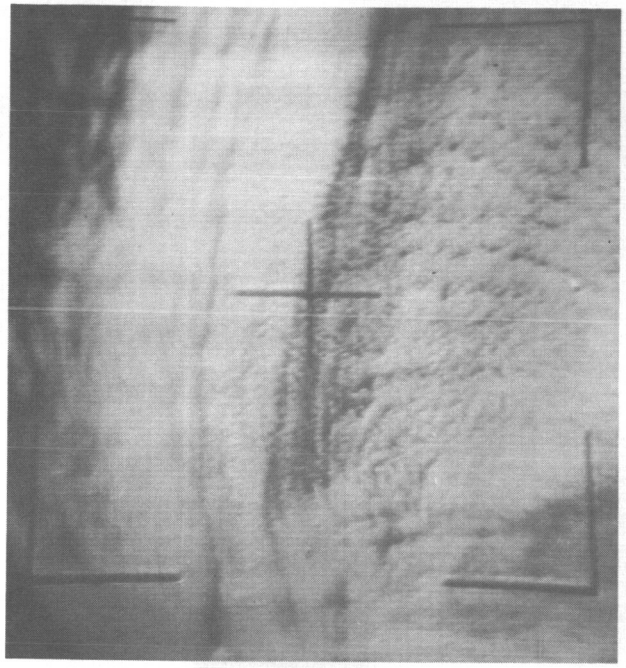


(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 8.146; bit compression ratio, 3.888.

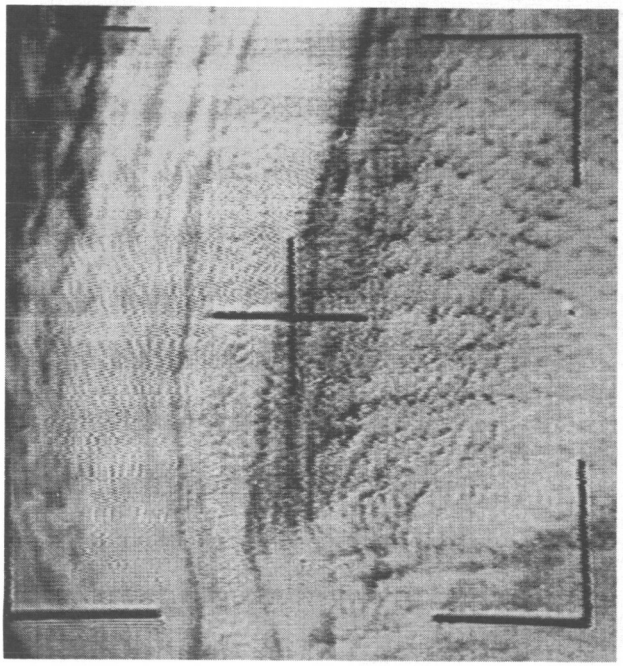


(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 6.413; bit compression ratio, 3.846.

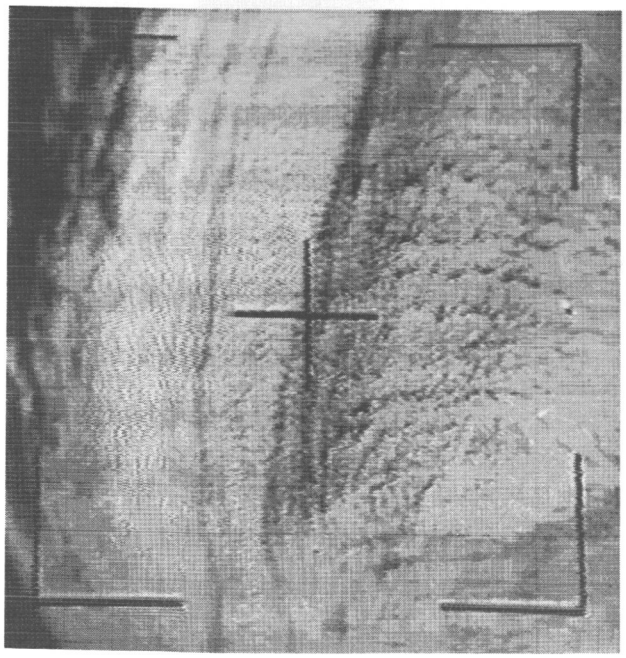
Figure 6-16—Pictures from Tiros III, orbit 102, frame 1, camera 1; taped before transmission from satellite; principal point, 11.5N, 4.0W; subsatellite point, 10.3N, 0.6W



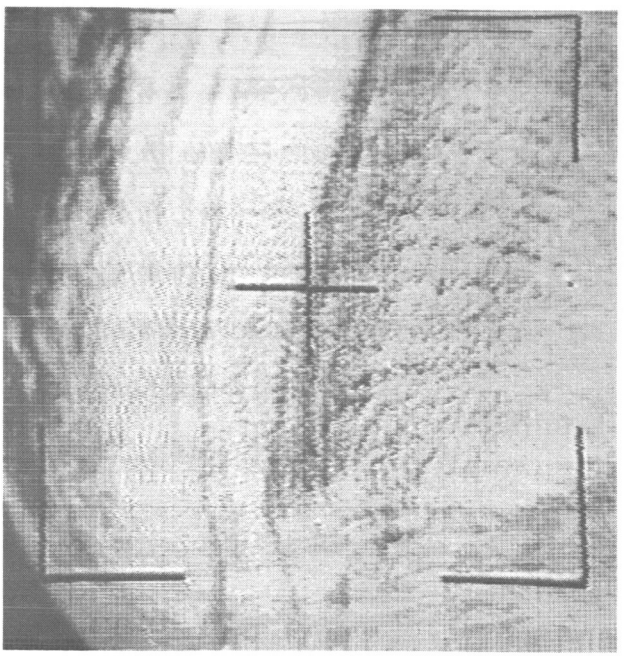
(a) Analog original.



(b) Digital original.

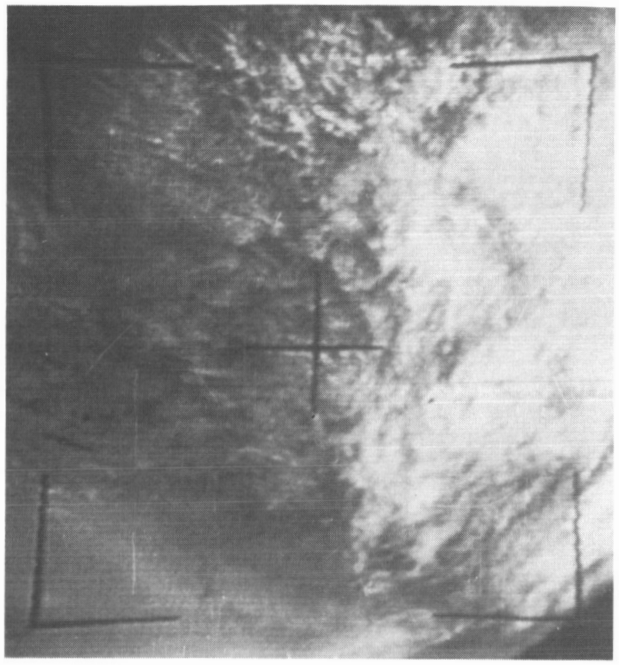


(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q = 4 bits per TV element; T = ± 1 level; L = 16 TV lines. Element compression ratio, 7.824; bit compression ratio, 3.762.

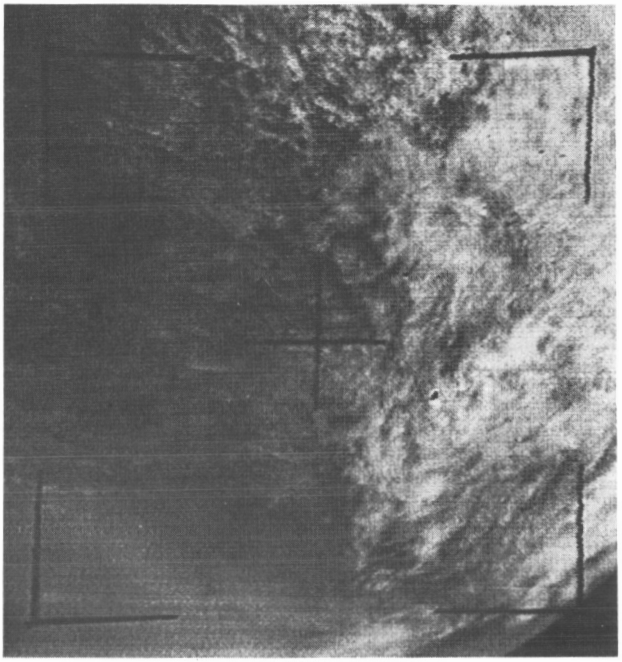


(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q = 6 bits per TV element; T = ± 2 levels; L = 16 TV lines. Element compression ratio, 6.863; bit compression ratio, 4.064.

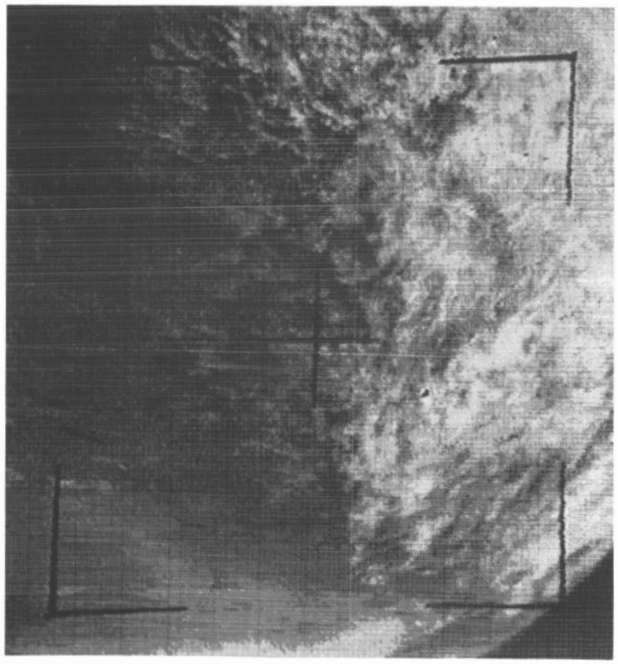
Figure 6-18—Pictures from Tiros V, orbit 3143, frame 6, camera 1; direct transmission from satellite; principal point, 32.4N, 69.3W; subsatellite point, 33.9N, 73.4W



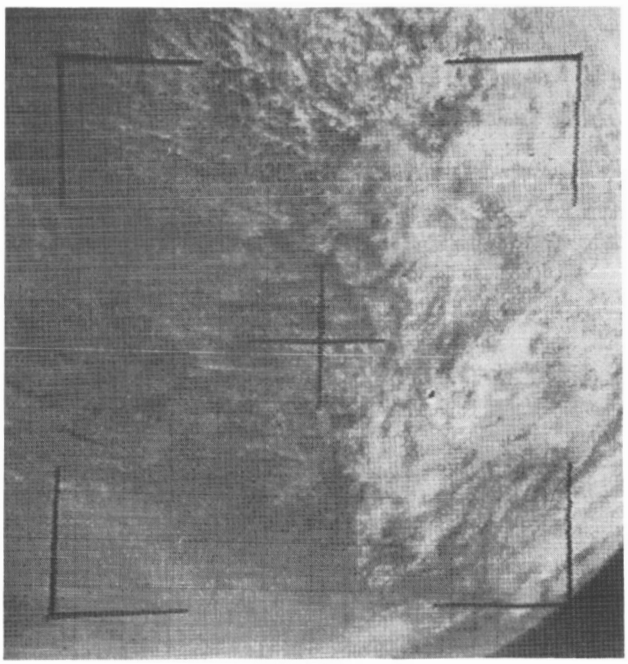
(a) Analog original.



(b) Digital original.

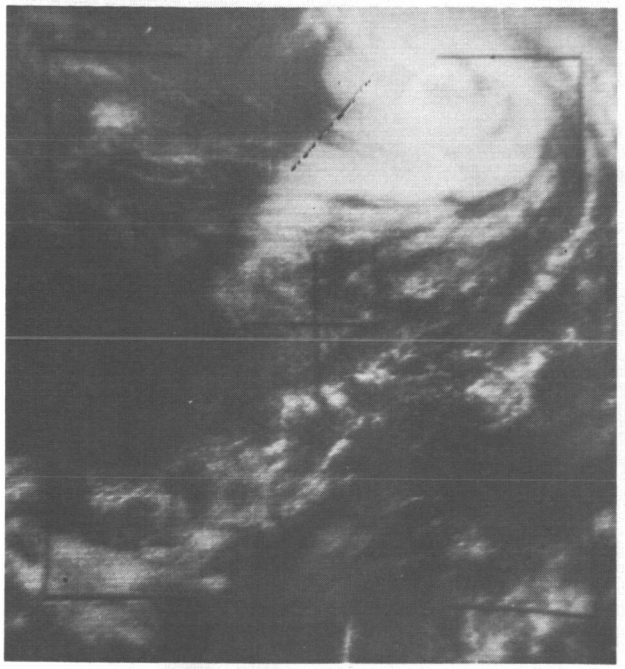


(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 7.379; bit compression ratio, 3.705.

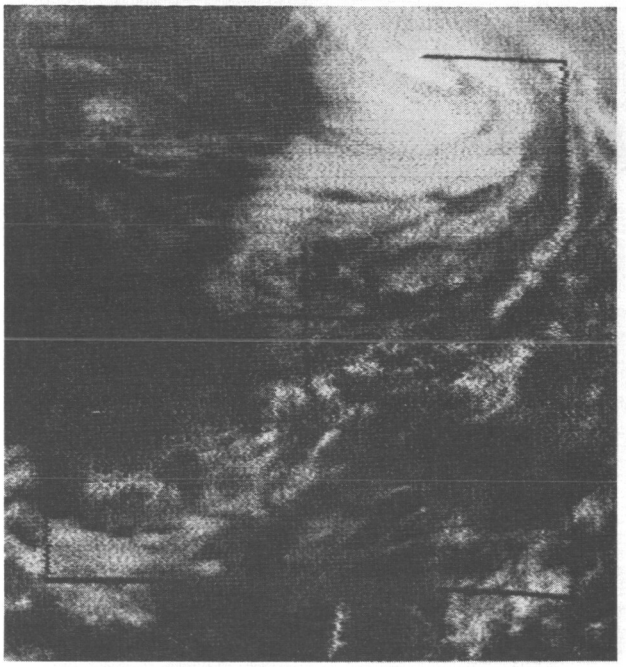


(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 5.913; bit compression ratio, 3.593.

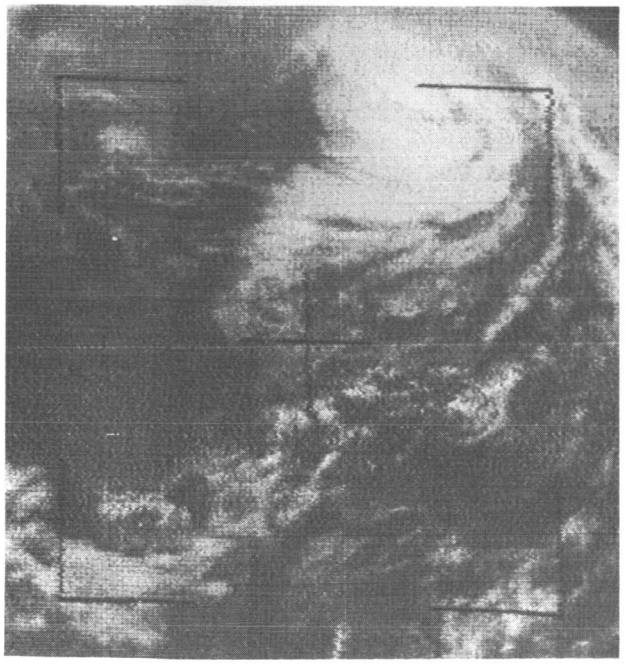
Figure 6-17—Pictures from Tiros III, orbit 102, frame 2, camera 1; taped before transmission from satellite; principal point, 13.3N, 5.6W; subsatellite point, 11.9N, 1.9W



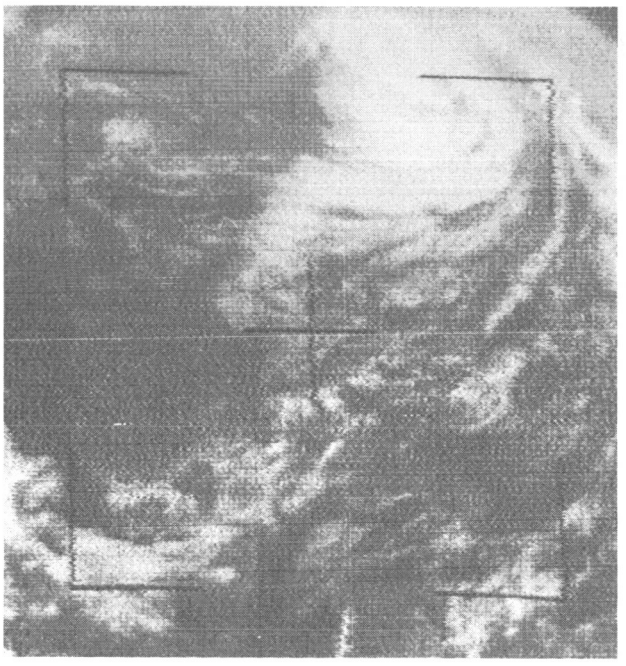
(a) Analog original.



(b) Digital original.

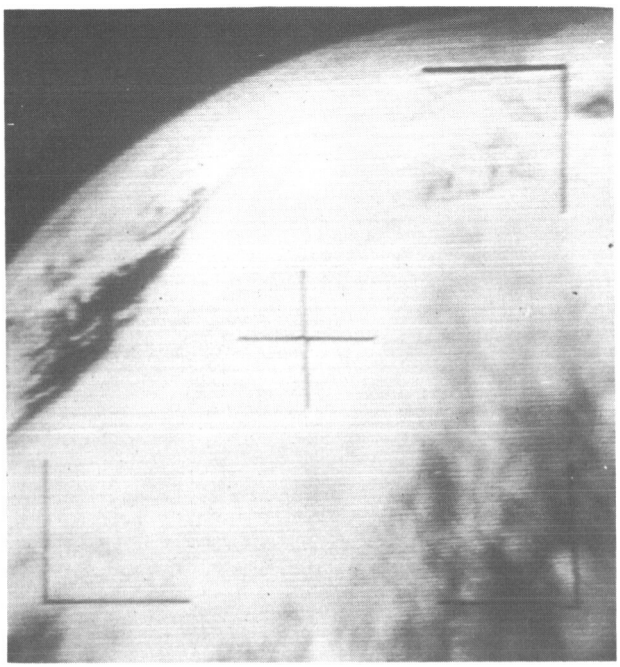


(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 3.890; bit compression ratio, 2.160.

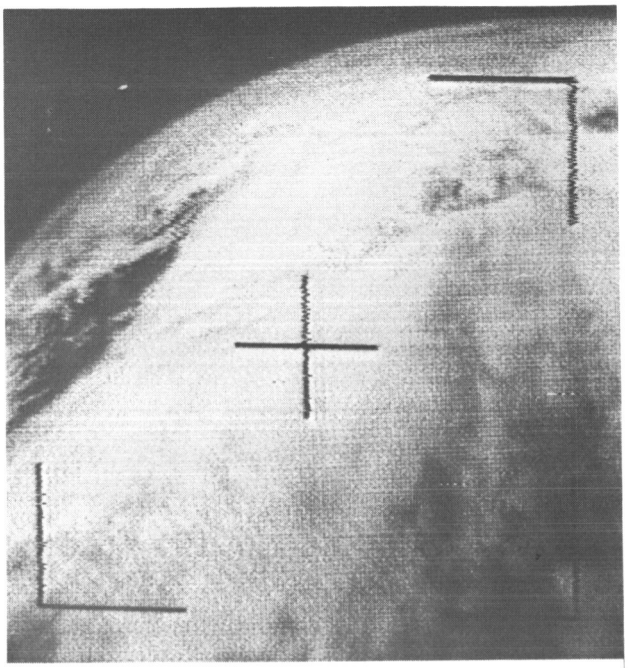


(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 3.297; bit compression ratio, 2.218.

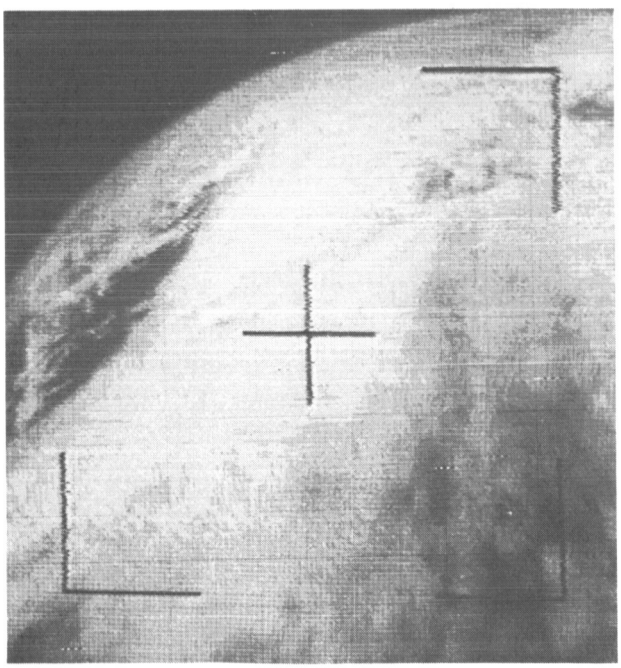
Figure 6-19—Pictures from Tiros VI, orbit 1100, frame 15, camera 1; direct transmission from satellite; principal point, 28.3N, 79.0W; subsatellite point, 26.1N, 79.8W



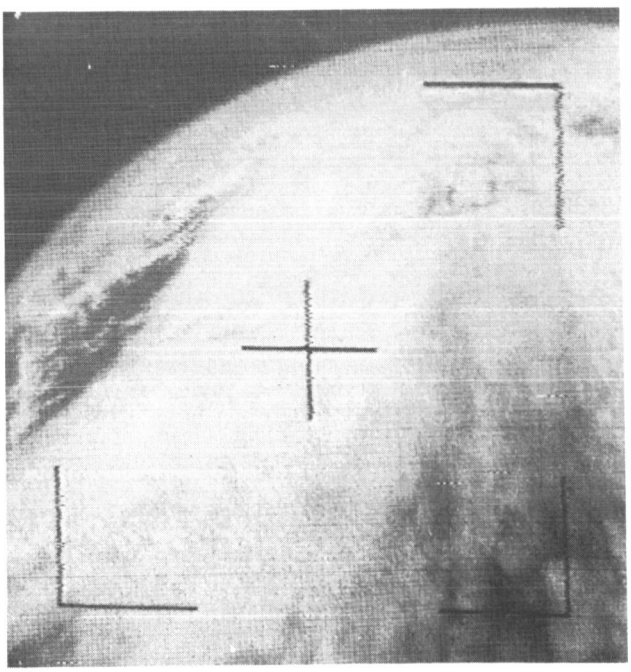
(a) Analog original.



(b) Digital original.

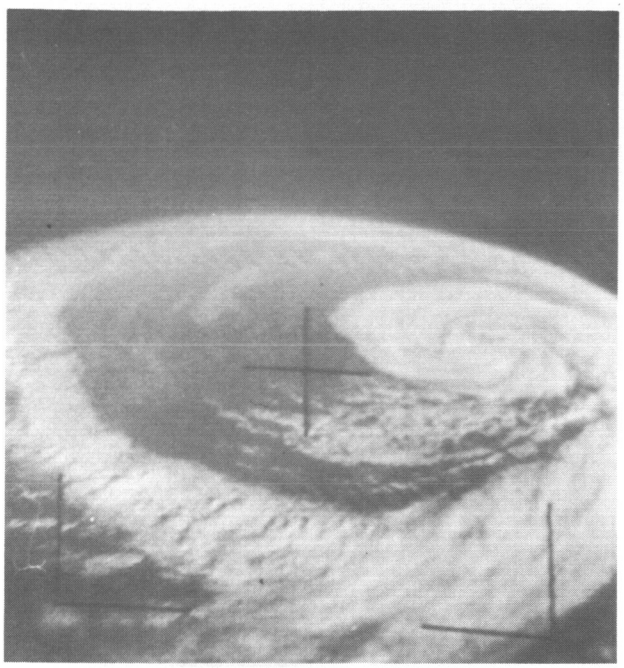


(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 9.690; bit compression ratio, 4.473.

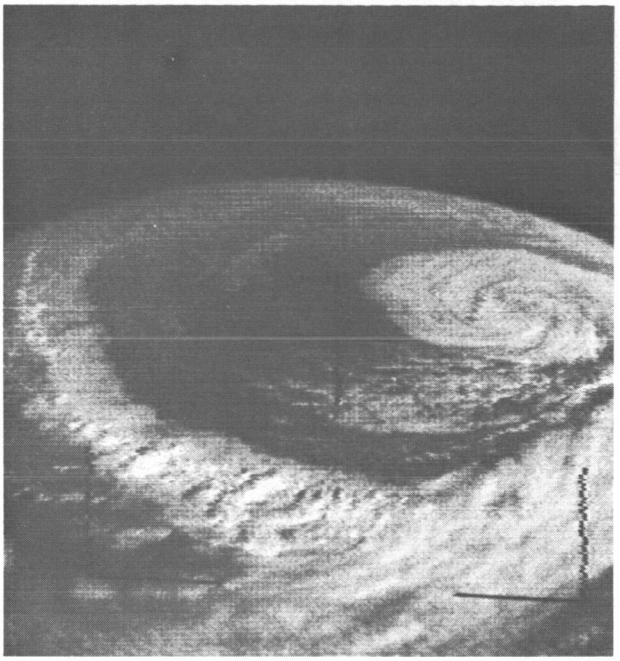


(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 7.875; bit compression ratio, 4.576.

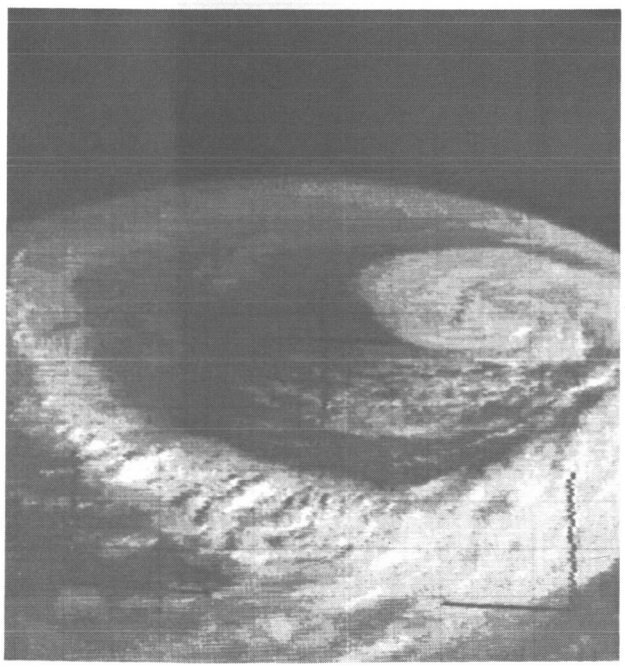
Figure 6-20—Pictures from Tiros VI, orbit 18, frame 21, camera 1; taped before transmission from satellite; principal point, 52.5N, 45.2W; subsatellite point, 50.4N, 37.3W



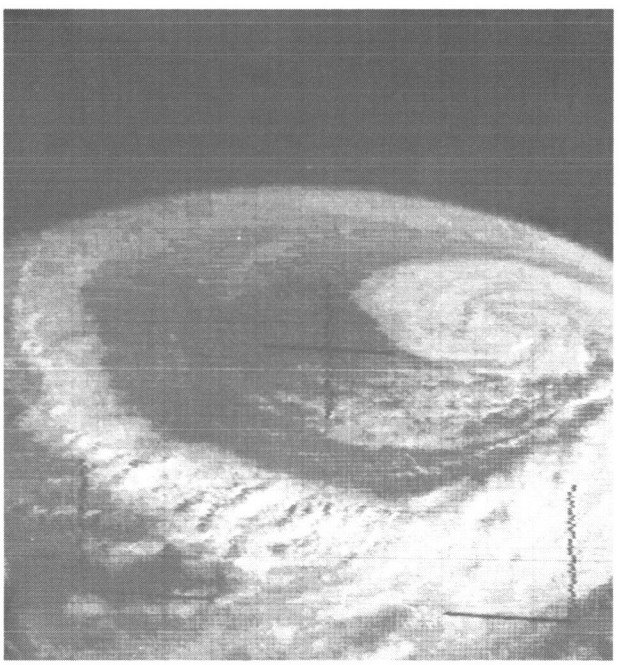
(a) Analog original.



(b) Digital original.



(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 19.954; bit compression ratio, 8.219.



(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 17.550; bit compression ratio, 9.127.

Figure 6-21—Pictures from Tiros VI, orbit 3692, frame 31, camera 1; taped before transmission from satellite; principal point, 36.8N, 57.2W; subsatellite point, 33.1N, 48.7W

to use data quantized to 5 bits per element and allow T = ±1, which is the same percentage error as 6 bits per element with T = ±2. Thus one would expect the element compression ratios for the 5-bit, T = ±1 case to be about the same as those for the 6-bit, T = ±2 case. If these 5-bit pictures were also displayed with 16 gray shades then they would possess about the same quality as the 6-bit pictures. The first-order entropies of the unmodified digital data quantized to 4, 5, and 6 bits per element are given in Table 6-5. The entropies for the 5- and 6-bit pictures are almost exactly the same, while the entropies for the 4-bit pictures are somewhat smaller.

Table 6-5-First-Order Entropies

Figure Number	First-Order Entropy for —		
	Q = 64 levels (6 bits/TV element)	Q = 32 levels (5 bits/TV element)	Q = 16 levels (4 bits/TV element)
12	4.510	4.472	3.512
13	4.736	4.622	3.647
14	4.777	4.678	3.701
15	4.728	4.656	3.689
16	4.257	4.201	3.227
17	4.606	4.566	3.578
18	4.516	4.449	3.483
19	4.649	4.561	3.590
20	4.891	4.813	3.829
21	4.478	4.443	3.532

The reader may find it interesting to compare the pictures processed by the conditional expectation predictor with those processed by the zero-order hold and linear predictors which are discussed in Reference 4. In general, the pictures processed by the conditional expectation predictor are of slightly better quality than zero-order-hold-predicted pictures. The pictures processed by the linear predictor are of higher quality than those processed by either of the other two methods.

APPLICATIONS FOR DATA COMPRESSION SYSTEMS

Some of the most obvious applications of data compression systems are in deep space communications, earth-orbiting operational spacecraft, and land-line data transmission. Figure 6-22 is a block diagram of both the transmitter and receiver ends of a data compression system model. At the transmitter end, the predictor accepts raw data from the information source. The predictor contains arithmetic, memory, and control functions which are arranged according to some prediction algorithm. Each raw data sample is compared with the corresponding predicted sample and the prediction error E_p is determined. If the prediction error exceeds some preset threshold T, the raw data sample must be transmitted in unmodified form. If, however, the prediction error is less than T, the sample is predictable and need not be transmitted. The comparator output is also fed back to the predictor to update the prediction mechanism. The encoder accepts raw unpredictable samples as well as indications of predictable samples and arranges this information according to some appropriate code. The information rate at the output of the encoder is certainly non-uniform. Since the main data-storage device would probably require a uniform read-in rate, a smoothing buffer is necessary.

At the receiver end, the decoder provides the predictor with all the data necessary to reconstruct the original message within the allowable prediction error. The predictor at the receiver is an exact copy of the predictor at the transmitter. After reconstruction, the message is transferred to the information sink.

The bit compression ratio can be a very useful parameter to a communications system designer. If C_B represents the bit compression ratio, then the designer can choose to reduce the transmission time to T/C_B for the same bandwidth or alternatively reduce the original bandwidth to B_W/C_B. If one desires to save power or reduce spacecraft weight by saving power, the signal power can be reduced by S/C_B without changing the S/N ratio, since the thermal noise is directly proportional to the bandwidth. In practice, however, one would probably choose to employ data compression techniques to achieve high communication channel efficiency. This can be achieved by keeping the information rate close to the channel capacity at all times. This implies a channel with the capability of adapting to the time-varying information rate.

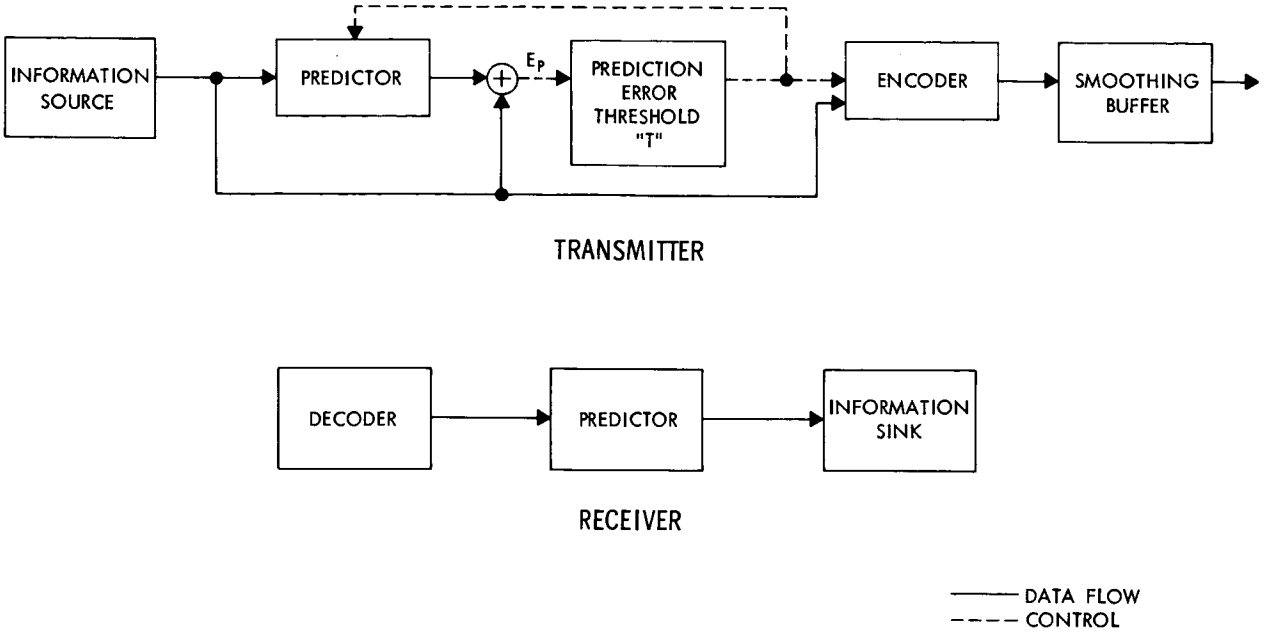


Figure 6-22–Block diagram of data compression system

CONCLUDING REMARKS

The most important outcome of this work was that the conditional expectation predictor produced compression ratios significantly greater than either the zero-order hold or the linear predictors. This result is true for each of the 10 TV frames used in the study and is significant since it shows that the conditional expectation predictor yields superior compression ratios despite the suspected nonstationary character of the information source. To summarize the numerical results, it should be noted that the conditional expectation predictor produced bit compression ratios (assuming ideal coding in the noiseless case) exceeding 5:1 on a number of single TV frames. It is also important that the cumulative compression ratio (10-picture average exceeded 4:1 for both the cases with 6 and those with 4 bits per TV element. At the same time the compressed pictures (Figures 6-12 to 6-21) seem to retain at least an acceptable level of quality.

Many interesting problems associated with adaptive data compression systems require further investigation. The prediction mechanism itself should be further developed to include, for example, the optimal relationship between learning period and memory size. Certainly the determination of efficient coding schemes for the adaptive data compression system for the noiseless channel is one of the most important problems still to be solved.

Further investigations might also consist in simulating noise environments for possible missions and analyzing effects on the noiseless-case coding structure in order to develop efficient error-correction codes. Investigations of this sort would eventually allow laboratory simulation of a complete compression system from the information source to the transmitter, through the communication channel to the receiver, and finally to the information sink. This arrangement would permit feasibility studies for specific missions as well as establish system design guidelines.

ACKNOWLEDGMENTS

Thanks are due to Mr. H. J. Gillis and Mr. O. N. Clark for programming of the IBM 7094 computer. The author is indebted to Dr. A. V. Balakrishnan of UCLA and Dr. L. D. Davisson of Princeton University for a number of interesting discussions and many helpful suggestions.

REFERENCES

1. Weber, D. R., "A Synopsis on Data Compression," Paper TA 1.1, Proc. 1965 National Telemetering Conf., (Houston, Texas, April 13-15, 1965), pp. 9-16.

2. Ellersick, F. W., "Data Compactors for Space Vehicles," Proc. Eighth Annual East Coast Conference on Aerospace and Navigational Electronics (Baltimore, Md., October 23-25, 1961), North Hollywood, Calif.: Western Periodicals Pub. Co., 1961, pp. 6.2.6-1 – 6.2.6-5.

3. Balakrishnan, A. V., "An Adaptive Nonlinear Data Predictor," Paper 6-5, Vol. II, Proc. National Telemetering Conf. (Washington, D. C., May 23-25, 1962), New York, N. Y.: The Institute of Radio Engineers, pp. 1-15.

4. Balakrishnan, A. V., Kutz, R. L., and Stampfl, R. A., "Adaptive Data Compression for Video Signals," NASA Technical Note D-3395, April 1966.

5. Davisson, L. D., "Data Compression and its Application to Video Signals," in: Final Report of the Goddard Space Flight Center Summer Workshop Program in the Analysis of Space Data and Measurement of Space Environments, GSFC Document X-100-65-407, June 15 to September 15, 1965, pp. A-35–A-55.

7. DATA COMPACTION STUDIES

C. R. Laughlin
Goddard Space Flight Center
Greenbelt, Maryland

The Systems Division of GSFC has been pursuing an active study program aimed at developing advanced techniques for reducing the channel capacity requirements for real-time digital transmission of imagery signals. Primary emphasis has been on studying the statistical nature of Tiros and Nimbus cloud cover photographs and on applying the modern mathematical theories of communication to these results. It is recognized, of course, that fundamentally the only way that a reduction in real-time channel capacity requirements can be effected is by simply not transmitting all of the original signal structure. However, modern information theory indicates that techniques exist (although most are yet to be developed) whereby a specific fidelity criterion can be attained although portions of the original signal structure are omitted; that is, the source output can be reduced without deterioration of the information content by means of redundancy removal.

Further discussion can be facilitated by referring to Figure 7-1 where a source encoder is shown. The operation of the source encoder is defined as the organization of the source signal in such a manner as to remove redundancy by means of future prediction of the signal samples. The source encoder forms a prediction on the basis of some amount of a priori knowledge of the statistical nature of the signal and upon complete knowledge of the past structure of the signal over some finite past history. The source decoder is assumed to operate in a corresponding manner so that (neglecting errors in the transmission channel), if a prediction of a future signal sample is found to be correct, that particular sample need not be transmitted to the receiver. If there exists a reasonable degree of correlation between signal samples (picture elements), then the prediction process will be successful most of the time and the total number of samples that must be sent will be greatly reduced.

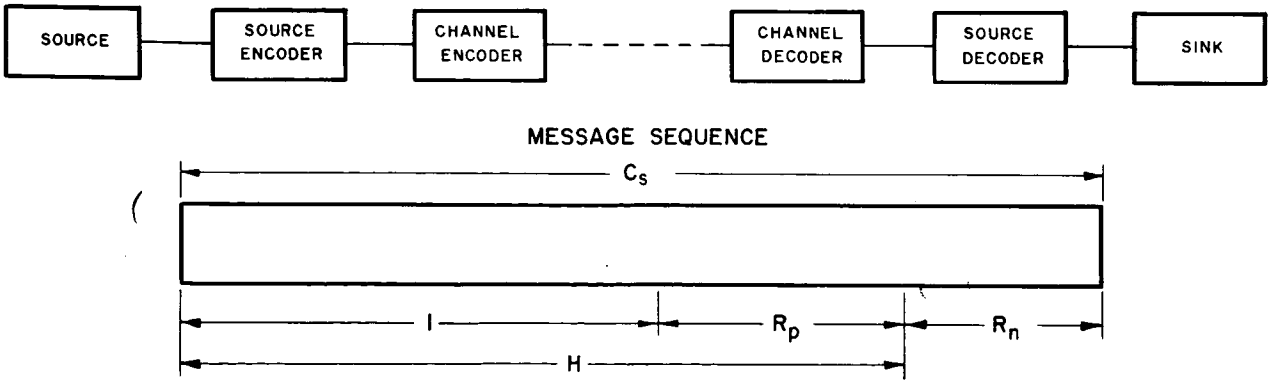


Figure 7-1—Diagram pertaining to transmission of information

The best predictor, in the information theory sense, is that operation which minimizes the entropy of the predictor's error signal. There are several approaches to the design of a predictor. For example, a linear Wiener predictor uses a weighted and summed set of previous signal values such that the mean square error is minimized. Alternatively, the predictor might employ the locally derived conditional probability distribution of the signal samples; in this case it is referred to as a Markov predictor.

In order to solidify the ideas of statistical redundancy in the information theory sense, one must turn to Figure 7-1 where the symbol C_s is used to designate the source capacity or the maximum information that the source could possibly deliver. The maximum value is attained when each of the possible levels ℓ occurs with equal probability, in which case $C_s = \log_2 \ell$ bits/element. The actual average information conveyed by each element in the sequence (or the entropy of the message) depends on the actual probability with which each of the ℓ levels occurs and is given by

$$H = - \sum_i p(i) \log_2 p(i) \quad \text{bits/element.} \tag{1}$$

The nonnegative quantity $(C_s - H)$ depends only on the first-order probability distribution through Equation 1 and is, therefore, independent of any sample-to-sample correlation. This is called nonpredictive redundancy. On the other hand, the average information per element I depends on the statistical influence extending over past elements so that if S_j denotes a state representing a particular one of the possible past sequences, I is given by

$$I = - \sum_{i,j} p(i, S_j) \log p_{S_j}(i) \quad \text{bits/element,} \tag{2}$$

where $p(i, S_j)$ is the joint probability that state S_j occurs and is followed by an element of the i^{th} level and p_{S_j} is the conditional probability that, given the occurrence of the state S_j, the next element will assume the i^{th} level.

The difference given by $(H - I)$ is called the predictive redundancy because it gives a quantitative measure of how well an element in the sequence can be predicted from a knowledge of past history and, therefore, provides a measure of how well a predictive encoding process can perform in a given situation. To complete the discussion of Figure 7-1, it is noted that the total statistical redundancy R is the difference $(C_s - I)$ between the source capacity and the actual information rate and can be written as the sum of the nonpredictive and the predictive redundancy;

$$R = C_s - I = (C - H) + (H - I) \quad \text{bits/element.} \tag{3}$$

The relative redundancy r will be useful in what follows and can be found by normalizing the total redundancy to the source capacity. That is,

$$r = \frac{R_n + R_p}{C_s} = 1 - \frac{I}{C_s} \quad \text{bits/element.} \tag{4}$$

Extensive investigation conducted by members of the Systems Division has revealed that element compression ratios of 3:1 or 4:1 can be readily attained on representative photographs. This ratio is taken as the ratio of the number of picture elements going into the source encoder (predictor) to the number of picture elements coming out, and does not account for any additional redundancy that may be added back in by the channel encoder. Such compression ratios do not come anywhere near the dreams and expectations of early investigators in the field and have even caused many investigators to drop their studies as being hopeless. A number of highly sophisticated approaches have been tried in the Systems Division by computer simulation and by fundamental statistical analysis. These include adaptive techniques optimized over local statistics. Results to date are in substantial agreement with those obtained by other organizations and with what has been reported in the open literature.

There is much evidence to support the contention that, for a large class of imagery signals, the autocorrelation function for small displacements is nearly exponential in behavior. Thus, linear prediction on the basis of previous elements alone will show little improvement when the

past history is extended over more than one or two previous elements. This implies that a linear, first-order Markov predictor (sometimes called a zero-order hold predictor) will perform nearly as well as the best device that can be devised. Our studies verify this and leave us with the conclusion that, on the whole, compression ratios on the order of two or three are the best that can be expected by exploiting the statistical properties of imagery signals with linear processing methods.

This conclusion is not altogether black because, while the gains are small, the zero-order hold predictor is relatively simple in terms of equipment complexity. An equally simple complete transmission system including such a predictor wherein this potential gain is fully realized remains to be devised. With such a system a compression ratio of 2:1 or 3:1 can be used for a corresponding savings in the transmission bandwidth, transmission time, or in the total energy used. For deep-space camera missions, this is clearly a worthwhile goal.

The first problem to be solved in this direction is to devise an efficient coding scheme for representing the compressed signal. It is hypothesized here that the predictor is a device which provides as its output a quantized signal which is one of ℓ levels and which represents the actual signal level during a sample interval whenever the prediction process fails. It is further assumed that, out of a video message of N elements in length, Q of the elements are predictable so that the inverse of the compression ratio given by

$$\frac{1}{K} = \frac{Q}{N} \tag{5}$$

approaches the limit in probability of the occurrence of a nonpredictable element which will be designated by q. In addition, when the prediction process fails, it is assumed that the predictor produces an additional signal (say of $\ell + 1$) for a total of $(N-Q)/N = 1 - (1/K)$ times. The limit in probability of the occurrence of this event will be designated as p.

On the basis of the above, the minimum number of bits per element N required to represent the predictor output which uses an ideal Shannon-Fano code can easily be shown to be

$$N = [p(\log p) + q(\log q)] + qL \quad \text{bits/element,} \tag{6}$$

where each of the signal levels is taken to be equally likely and L equals $\log_2 \ell$. The first two terms of Equation 6 within the brackets assumes a value between zero and one for all admissible values of p and q so that N is within the bounds

$$qL \leq N \leq 1 + qL. \tag{7}$$

Expression 7 states the encoding problem in concise form. For example, with a six-bit-per-element picture signal which can be compressed by a 3:1 ratio (i.e., $K = N/Q = 3 \approx 1/q$), ideally no less than two, nor more than three, bits per element are required to encode the predictor output.

Early attempts to encode compressed data were based on run-length encoding schemes which depend upon predictable elements occurring in groups or runs. Studies in the Systems Division have resulted in a different approach which actually achieves the upper bound of Expression 7 in all cases. This coding scheme is indicated in Figure 7-2, the timing diagram is shown in Figure 7-3 and the logic flow diagram, in Figure 7-4.

The essential feature of this system is a highly efficient coding scheme which is practical to instrument. A digital-scan camera is required which effectively produces a variable rate source and thereby reduces buffer requirements. The camera scans a line at a time at an element-per-second rate equal to the bit transmission rate. The predictor also operates at this rate and stores the value of each unpredictable element in the Q store. At the same time, the T-register is loaded

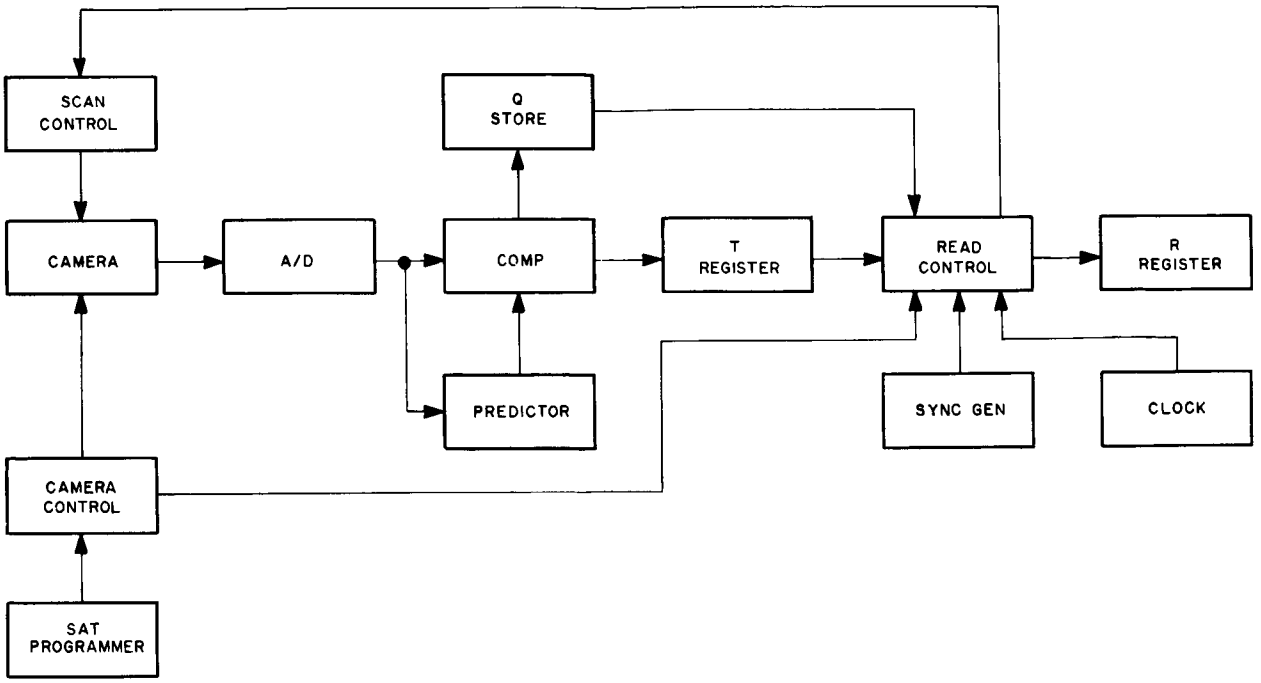


Figure 7-2—Functional diagram

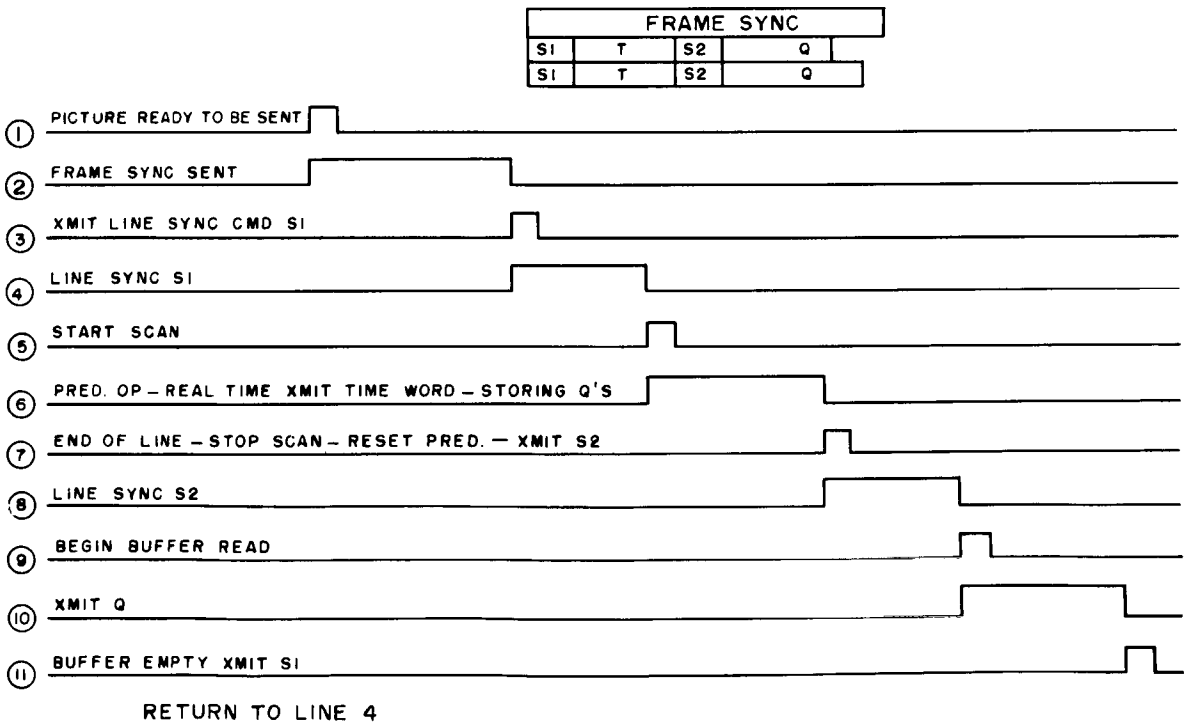


Figure 7-3—Timing diagram

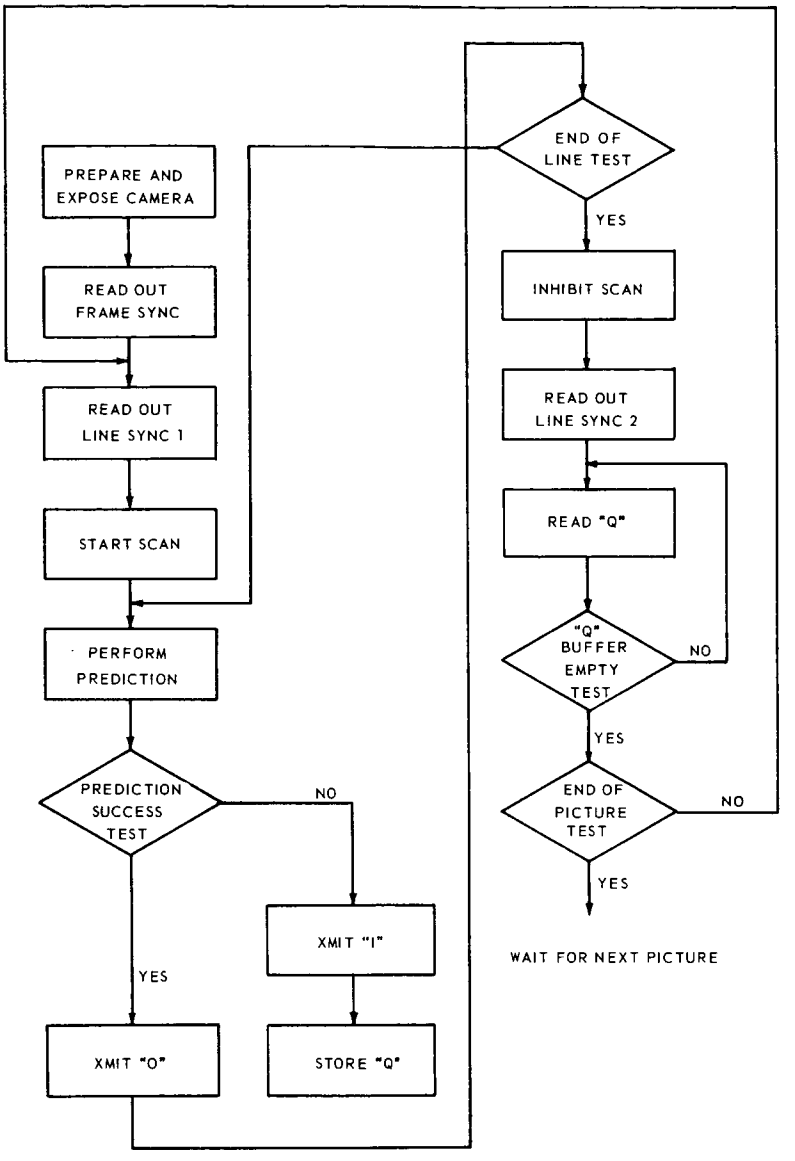


Figure 7-4—Logic diagram

with a sequence of prediction success indicator digits which represent the position of each unpredictable (or predictable) element in the line. This sequence is fed directly to the R-register for transmission to the channel. At the completion of the T readout, the camera scan is stopped at the end of the given line while the Q store is fed to the R-register for transmission. The readout control is an electronic commutator which formats the output sequence.

The efficiency of the proposed coding scheme is indicated in Table 7-1 for ten representative photographs. The performance of the best run-length coding scheme is also presented in this table. It is expected that certain portions of the proposed system will be fabricated and optimum synchronization methods will be studied.

Table 7-1–Comparison of Run-Length Encoding and Position Encoding
(Data from computer analysis of zero-order hold predictor
operating on 10 Tiros pictures)

Picture	Bit Compression Ratio		Bits per Element	
	Run-Length Encoding	Position Encoding	Run-Length Encoding	Position Encoding
1	2.473	2.602	2.097	1.922
2	2.237	2.469	2.235	2.025
3	2.105	2.381	2.376	2.099
4	2.263	2.481	2.209	2.015
5	1.762	2.129	2.838	2.349
6	1.714	2.092	2.916	2.390
7	2.127	2.349	2.351	2.128
8	1.216	1.568	4.112	3.188
9	1.943	2.267	2.573	2.206
10	4.090	3.276	1.222	1.526
Average	2.006	2.288	2.493	2.185

Session III
ADAPTIVE ENCODING DECODING, AND MODULATION

Chairmen G. Ludwig
GSFC

R. Rochelle
GSFC

Contributors J. Snively, Jr.
GSFC

R. Cliff
GSFC

C. Creveling
GSFC

8. A PARAMETER EXTRACTION TECHNIQUE

J. W. Snively, Jr.
Goddard Space Flight Center
Greenbelt, Maryland

INTRODUCTION

A so-called "parameter extraction technique" which is applicable to data displayed in histogram form is presented herein. The method is called a parameter extraction technique rather than a data compression technique because not all of the information contained in the original histogram is preserved. To paraphrase a remark made by Dr. Balakrishnan, the parameters sent are tailored to the ultimate use of the data.

The quantiler technique described in a preceding paper of this symposium* is another example of a parameter extraction technique. However, the technique discussed in the present paper is more algebraic than the statistical approach given in the preceding paper since calculated upper and lower bounds on the magnitude of the desired quantities are used rather than confidence levels or hypothesis tests.

The technique described in this paper has been implemented into flight hardware as part of the GSFC/University of Maryland plasma experiment to be flown on Interplanetary Monitoring Platform (IMP) flights F and G (Reference 1). The IMP's are small spin-stabilized spacecraft launched into highly elliptical Earth orbits. In this experiment, the satellite spin is divided into 16 equal sectors and the particles arriving at the sensor in each sector are counted; thus a histogram is produced. Figure 8-1 shows a pair of typical histograms. Histograms which resemble a very sharp normal distribution are expected from interplanetary space outside the magnetosphere, whereas in the transition region just inside the shock wave that surrounds the magnetosphere, a relatively flat curve is expected. The goals of this experiment were to map the boundaries of the transition region and the magnetosphere and at the same time to study relatively fast changes in the plasma intensity. These goals could not be achieved within the bandwidth allotted to the experiment.

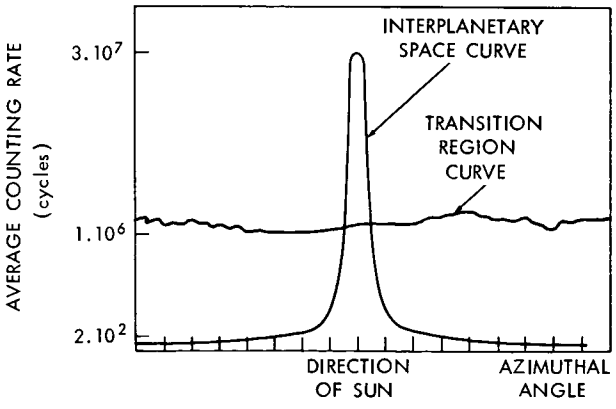


Figure 8-1—Typical histograms expected in a plasma-measuring experiment

In IMP flights F and G (this constitutes this parameter extraction technique) the area of the histogram and the sum of squares of the histogram bars are to be calculated. The present paper elaborates on this technique and shows how well certain properties of the original histogram can be recovered from these parameters.**

*Paper 4, pp. 55 to 73.

**Statements made throughout this paper are supported by mathematical proofs in the Appendix of Reference 2.

TECHNIQUE

Relationship between Area and Sum of Squares for a Histogram

Assume that the area A of a histogram is known. The sum of the squares of the histogram bars Σ must lie between two numbers which differ by a factor equal to the number of bars n in the histogram. The larger of these two limits for the sum of the squares corresponds to a histogram where all but one of the bars contain no counts. Therefore, this larger limit for the sum of the squares is simply the square of the area, A^2 . The smaller limit for the sum of the squares corresponds to a histogram where all of the bars have equal numbers of counts. In this case, the lower limit turns out to be A^2/n .

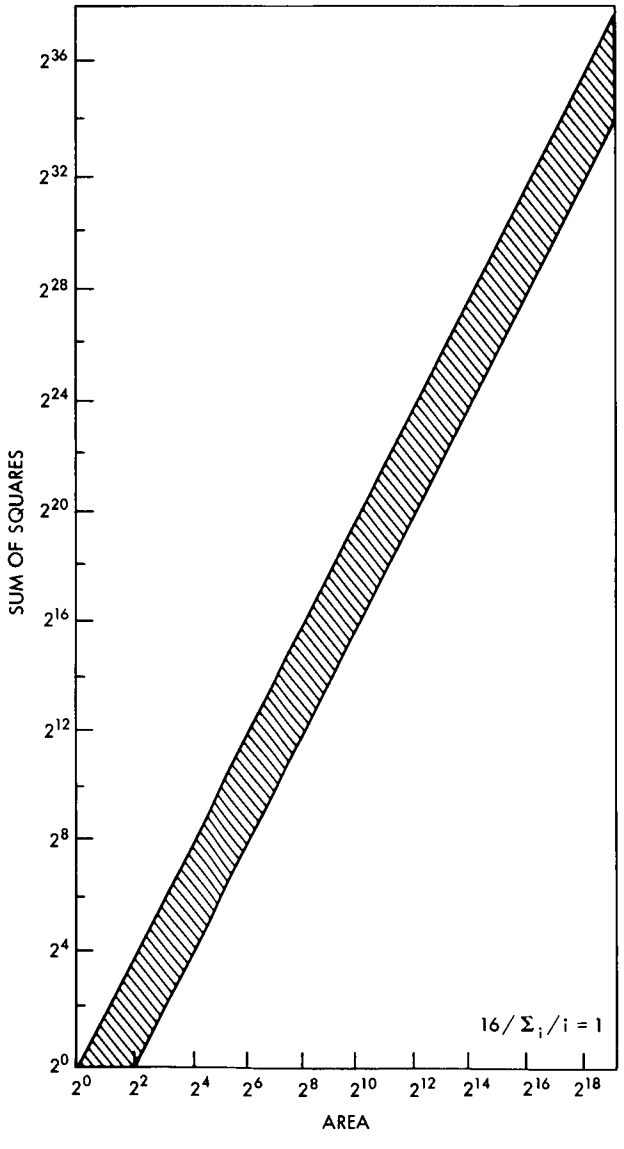


Figure 8-2—Relationship between area and sum of squares for a histogram with 16 bars

These facts are illustrated in Figure 8-2 which shows the relationship between the area and the sum of the squares for histograms of 16 bars. Histograms having 16 bars (with area less than 2^{19} counts) lie within the cross-hatched region between the two parallel lines. For example, a 16-bar histogram with an area of 2^6 must have its sum of squares between 2^{12} and 2^8 . This absolute constraint is the basic fact taken advantage of in the telemetry technique we wish to describe.

Hardware Use of This Relationship

If the quantity A is known, then, according to the preceding section, the sum of squares Σ is known to within a factor of n. In other words, the location of the most significant bit of Σ is known to within $\log_2 n$ bit positions. Specifically, for a 16-bar histogram the most significant bit must lie within one of four positions. If for this 16-bar histogram A is 2^6 counts, then the most significant bit of Σ is either the 9th, 10th, 11th or 12th bit of the word. The telemetry technique being expounded is simply the transmission of A and $\log_2 n$ bits of Σ. More bits of Σ can be transmitted if more accuracy is desired. Figure 8-3 is a block diagram of the equipment designed to implement this technique for the IMP plasma experiment.

The area counter at the bottom of the diagram commutates four bits of the sum of the squares counter to the telemetry system. Note that in this application a logarithmic counter (Reference 3) is used for the area determination. Therefore, although the total area of the collected histogram can be as large as 2^{19} counts, only eight bits are required to represent this number to a ±3 percent accuracy.

Transmission of the 12 indicated bits allows one to determine the sum of the squares to, at worst, ±33 percent. This greatest error occurs when the four commutated bits of the sum of the squares counter are 0001. When these four bits are 1111 the worst error is ±3.3 percent.

The remainder of this paper is theoretical. It assumes that the area and the sum of the squares are known exactly and consequences of such conditions will be discussed. The consequences obtained may be applied to cases where the area and the sum of the squares are not known exactly.

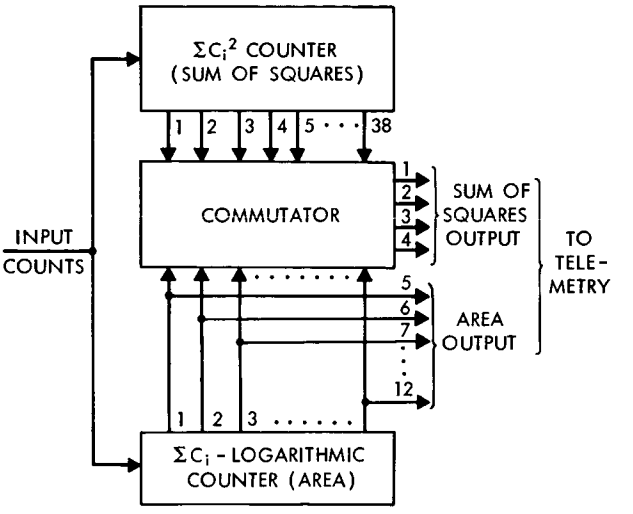


Figure 8-3—Block diagram of processing equipment

INFORMATION CONTENT OF THE AREA AND THE SUM OF SQUARES

Definition of Ratio r

A quantity denoted by the symbol r which we shall call the "ratio" is defined as

$$r = \frac{n \sum C_i^2}{\left(\sum C_i\right)^2} \quad . \tag{1}$$

Here n denotes the number of bars in the histogram and C_i denotes the number of counts in the i^{th} largest bar. This parameter is related to the mean μ and the variance σ^2 by the equation

$$r = 1 + \frac{\sigma^2}{\mu^2} \tag{2}$$

but has the advantage of assuming only values between 1 and n. For this reason, it is a useful parameter for describing various properties of histograms.

Peak Height

The value of the largest bar of a histogram, expressed as a fraction of the area of the histogram, cannot be greater than the area A nor smaller than an n^{th} of the area A/n, where n is the number of bars in the histogram. If the largest bar has the value A then all other bars must have the value zero. If this largest bar has the value A/n all others must have this same value. These two cases correspond to ratios of n and of 1, respectively. These are the extreme cases.

For any given ratio, the largest bar of a histogram can assume only values in a subinterval of the interval between A/n and A. For a given ratio r the greatest possible value for the largest histogram bar is given by

$$\frac{A}{n}\left(1 + \sqrt{(n-1)(r-1)}\right). \tag{3}$$

This expression reduces to A when $r = n$ and to A/n when $r = 1$. Therefore, the expression for the largest histogram bar agrees with the known histograms for the extreme values of r.

If the largest bar of a histogram with ratio r is equal to the value given by Expression 3, that histogram in this special case is completely determined. In fact, all of the remaining $n - 1$ bars are equal. Therefore, these bars must be equal to the value of the expression

$$\frac{a}{n}\left(1 - \sqrt{\frac{r-1}{n-1}}\right), \tag{4}$$

which is merely the area not included in the large bar divided by $n - 1$. Figure 8-4 shows several 5-bar histograms. Each histogram has the greatest largest bar possible for the indicated r. Note that the largest bar decreases as r decreases and the base of the histogram increases as r decreases.

The ratio r of any histogram must lie in the interval between 1 and n. Divide this interval into $n - 1$ subintervals (or S-intervals) and label these intervals from 2 to n as shown in Figure 8-5. Any subinterval S contains ratios between n/S and $n/(S - 1)$. This integer S is the fewest number of nonzero bars possible for histograms with ratios in the S-interval. For example, if $r = 1$ then $S = n$ since 1 lies in the interval bounded by 1 and $n/n - 1$. Therefore, n bars must have nonzero values for a ratio of 1. Another example is if $n = 16$ and $r = 5$ then, since 5 lies between 16/4 and 16/3, $S = 4$. Therefore this 16-bar histogram must have at least 4 nonzero bars. There is no restriction as to how many nonzero bars a histogram may have.

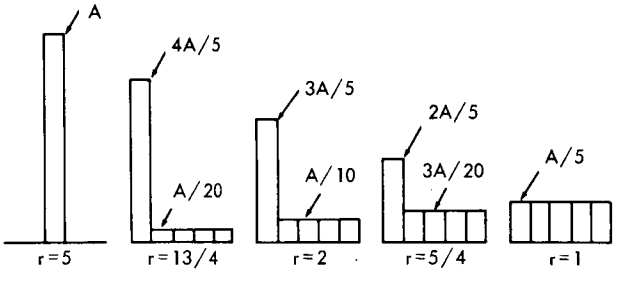


Figure 8-4—Histograms (n = 5) with the largest peak height for the indicated ratios

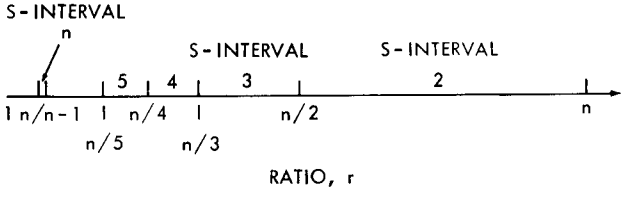


Figure 8-5—A segment of the real number line illustrating the definition of s-intervals

The smallest possible value for the largest bar is

$$\frac{A}{S}\left(1 + \sqrt{\frac{rS - n}{n(S-1)}}\right). \tag{5}$$

When $r = n$ we have $S = 2$, and the expression reduces to A. When $r = 1$ we have $S = n$ and the expression reduces to A/n. Therefore, this expression agrees with the known histograms for the extreme values of r.

If the largest bar of a histogram with ratio r is equal to the value given by Expression 5, that histogram in this special case is completely determined. It is a histogram with S bars, $S - 1$ of which are equal to the value of Expression 5. The remaining bar has the value

$$\frac{A}{S} = \left(1 - \sqrt{\frac{(S-1)(rS-n)}{n}}\right). \tag{6}$$

Figure 8-6 shows several 5-bar histograms. Each histogram has the smallest largest bar possible for the indicated r. Note that the largest bars decrease as r decreases and that the smaller bar increases as r decreases until it equals the largest bars. Then as r continues to decrease a new bar increases from zero until it equals the larger bars again. This process continues until we have n equal bars.

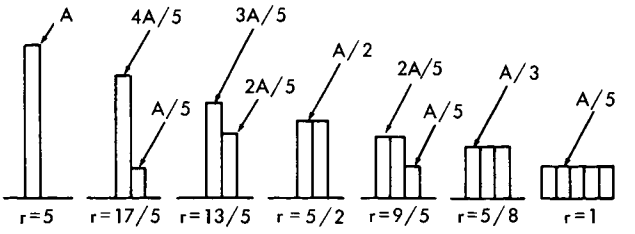


Figure 8-6—Histograms (n = s) with the smallest peak height for indicated ratios

Figure 8-7 illustrates, for 16-bar histograms, how the bounds on the largest histogram bars must lie within the crosshatched region between the two bounds.

Figure 8-8 is a plot for 16-bar histograms of the largest possible plus-or-minus percentage error in determining the largest histogram bar as a function of the ratio. Note that for the sharp peaked curves the percentage error is quite small. The maximum percentage error of ±43 percent occurs for a histogram with a ratio of slightly less than 2. This corresponds to a flat curve where the peak height is of lesser significance than it is for the very peaked curves of the higher ratios.

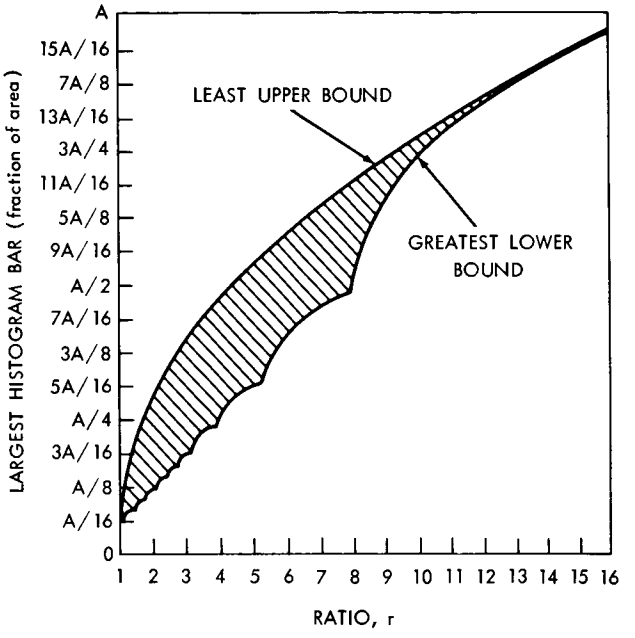


Figure 8-7—Bounds for the largest histogram bar (n = 16)

Amplitude Distribution

Another quantity which can be deduced from a knowledge of the area and the sum of the squares of a histogram is amplitude distribution. In fact, one can easily obtain bounds for each bar of a histogram that are similar to those for the largest one. Let C_1 refer to the largest bar of a histogram, C_2 to the second largest bar, and so forth, so that C_n refers to the n^{th} largest bar of the histogram. In this notation for the case of a histogram with n bars, C_n will be the smallest bar of the histogram.

The largest possible value for the p^{th} largest bar of a histogram with area A and ratio r must be computed in one of two ways depending on the value of the ratio. If the ratio is between n/p and n, the largest possible value for the p^{th} largest bar is given by

$$\frac{A}{p}\left(1 - \sqrt{\frac{rp - n}{n(p - 1)}}\right). \qquad (7)$$

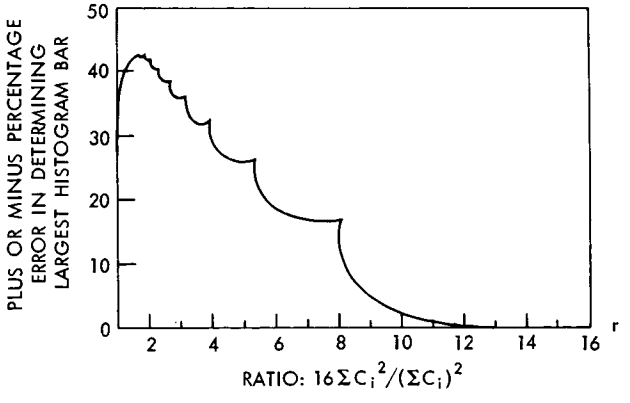


Figure 8-8—Percentage error in the largest histogram bar vs. ratio

If the ratio is between 1 and n/p, the largest possible value for the p^{th} largest bar is given by

$$\frac{A}{N}\left(1+\sqrt{\frac{(n-p)(r-1)}{p}}\right) \quad (8)$$

If $p = 1$, the entire range of ratios is covered by Expression 8, which reduces to Expression 3 when 1 is substituted for p.

When the ratio is equal to n/p (except for $p = 1$) both Expressions 7 and 8 give identical results, namely, that the largest possible value for the p^{th} largest histogram bar is exactly A/p. This value is the largest possible value that the p^{th} largest bar of a histogram can have for any ratio. It corresponds to a histogram with exactly p, equal, nonzero bars.

If the ratio is between 1 and $n/(p-1)$, the smallest possible value for the p^{th} largest histogram bar ($p \geq 2$) is

$$\frac{A}{n}\left(1-\sqrt{\frac{(p-1)(r-1)}{n-p+1}}\right) \quad (9)$$

When the ratio is 1, Expression 9 reduces to A/n. This corresponds to the known extreme histogram for this ratio, namely the one with n equal bars. When the ratio is $n(p-1)$, Expression 9 reduces to zero.

Figure 8-9 illustrates how the bounds on some of the histogram bars vary with ratio for the case where $n = 16$. Note that bounds on the largest bar were already illustrated in Figure 8-7.

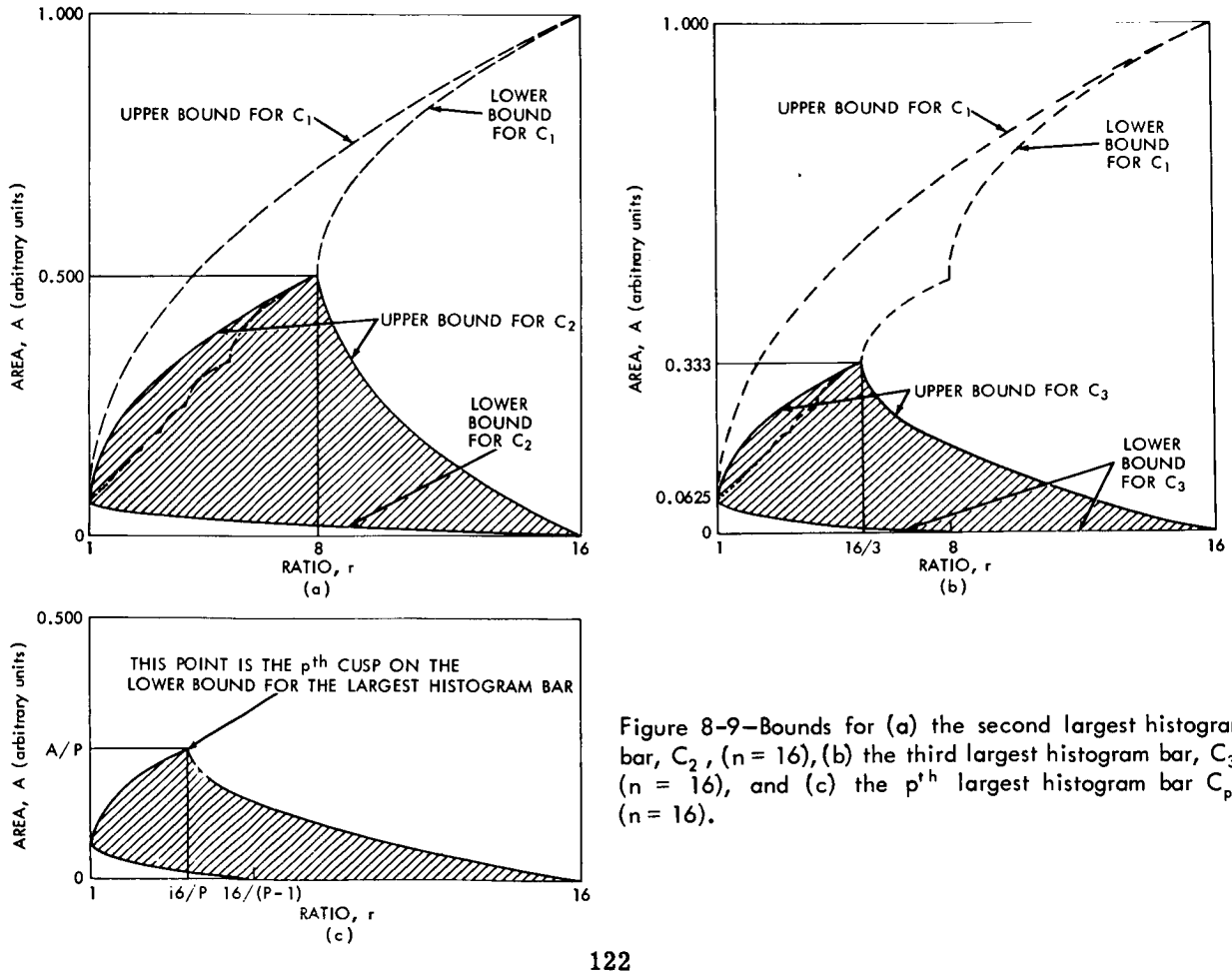


Figure 8-9—Bounds for (a) the second largest histogram bar, C_2, ($n = 16$), (b) the third largest histogram bar, C_3, ($n = 16$), and (c) the p^{th} largest histogram bar C_p, ($n = 16$).

For bars other than the largest, the least upper bound for the p^{th} largest bar increases from A/n when $r = 1$ to A/p when $r = n/p$, and then decreases to zero when $r = n$. The greatest lower bound decreases from A/n when $r = 1$ to zero when $r = n/(p - 1)$. The two extreme cases $r = 1$ and $r = n$ are the only completely determined ones.

Figure 8-10 illustrates the consequences of these results for the case where $n = 16$. For ratios with integer values, the bounds for each histogram bar are drawn with the largest bar on the left and the smallest bar on the right. In this form one can see at a glance how the shape of the resulting histograms varies as the ratio changes. Note that histograms with larger ratios are much steeper and narrower than those with lower ratios.

FURTHER HARDWARE CONSIDERATIONS

In the preceding section the information content of the area and sum of squares of a histogram was discussed. The discussion was based on these quantities' being known exactly. In practice, however, as in the IMP plasma experiment, these quantities will not be known exactly, but for a given set of bits these quantities will be known to lie between certain known bounds. For example, suppose the area A is known to lie between A_{max} and A_{min} and the sum of squares Σ is known to lie between Σ_{max} and Σ_{min}. Then the ratio r must satisfy the bounds

$$\frac{n \sum_{min}}{A_{max}^2} \leq r \leq \frac{n \sum_{max}}{A_{min}^2} \tag{10}$$

Similarly, bounds for any histogram bar could be found by choosing the largest upper bound and the smallest lower bound from among those for all ratios in the range of Expression 10.

When one uses the flight hardware for the IMP plasma experiment designed for a maximum number of counts in any histogram bar of 2^{17} and a maximum total area over 16 bars of 2^{19} counts, the ratio r is always determined to better than ±40.0 percent of its range, but on the average it is determined to about ±12.0 percent.

Table 8-1 is a segment of a computer program output which relates the output of the IMP flight hardware to the input data which produced the specified output. For example, if the log counter is 189 (275 octal), the area of the input histogram lies between 29,696 counts and 30,719 counts. If, furthermore, the squarer output is 12, the ratio of the input histogram must be between 13.64 and 15.85 and the largest bar of this input histogram is between 27,312 counts and 30,571 counts. For each of these quantities the harmonic mean (H. M.) of the range and the maximum plus or minus percentage error (P.E.) are also listed.

CONCLUDING REMARKS

The present paper has shown how transmission of the area of a histogram and certain bits of the sum of the squares enables one to recover much of the original histogram. It has shown how knowledge of the area of the histogram imples a restriction on the sum of the squares of the histogram. This fact is the backbone of the entire process for it enables the sum of the squares to be transmitted with a smaller number of bits if the area is also transmitted. This paper has also shown how knowledge of both the area and the sum of the squares of a histogram implies the restrictions on each of the bars of the histogram.

As spacecraft travel farther away from Earth, the need for onboard processing of data will increase. The IMP plasma experiment is an example of a situation where onboard processing is necessary so that the desired goals can be accomplished. The computation discussed in this paper

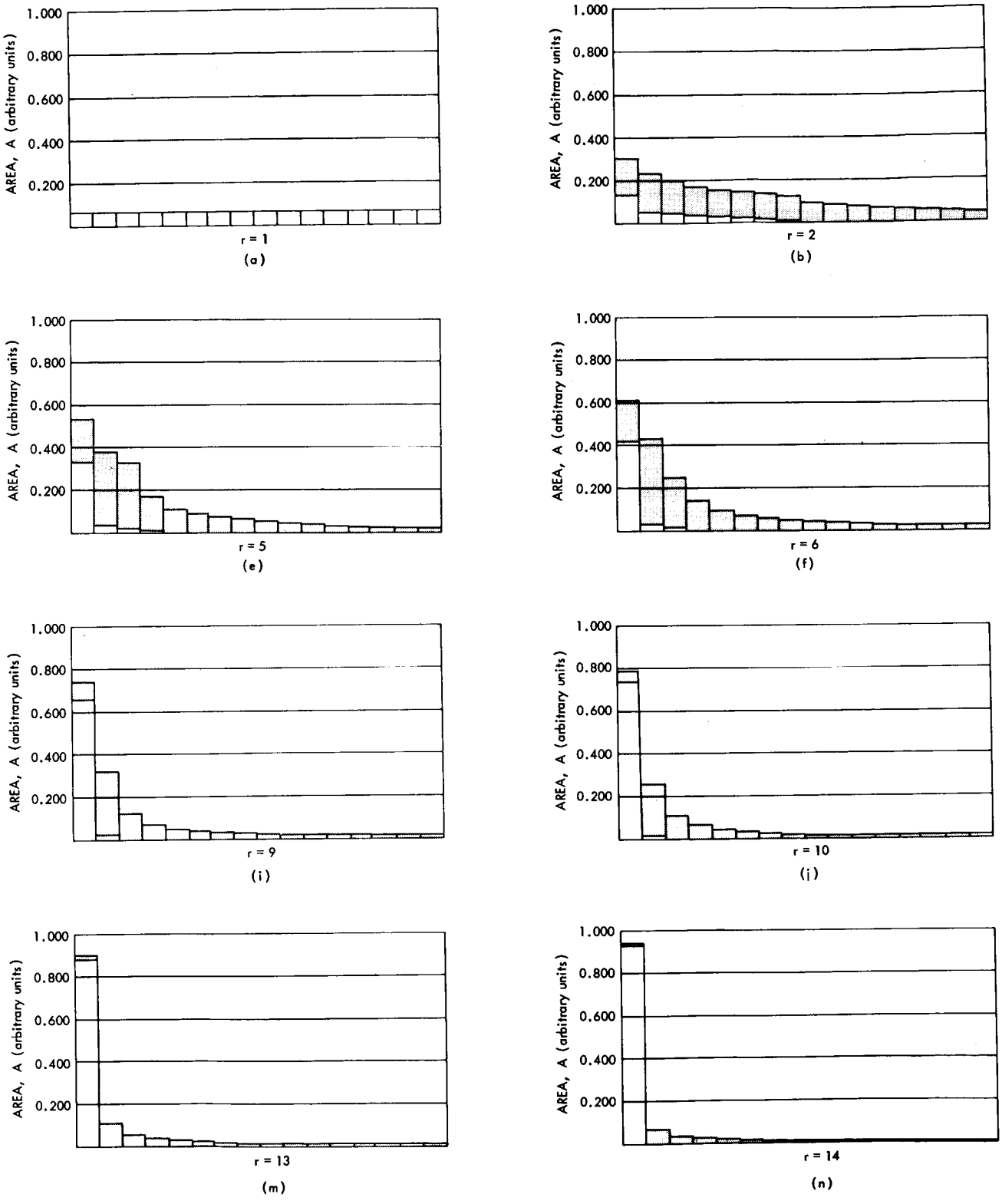


Figure 8-10(a through p)– Largest and smallest bars for histograms with the indicated ratios. (The bars of any histogram with the indicated ratio must lie in the shaded region.)

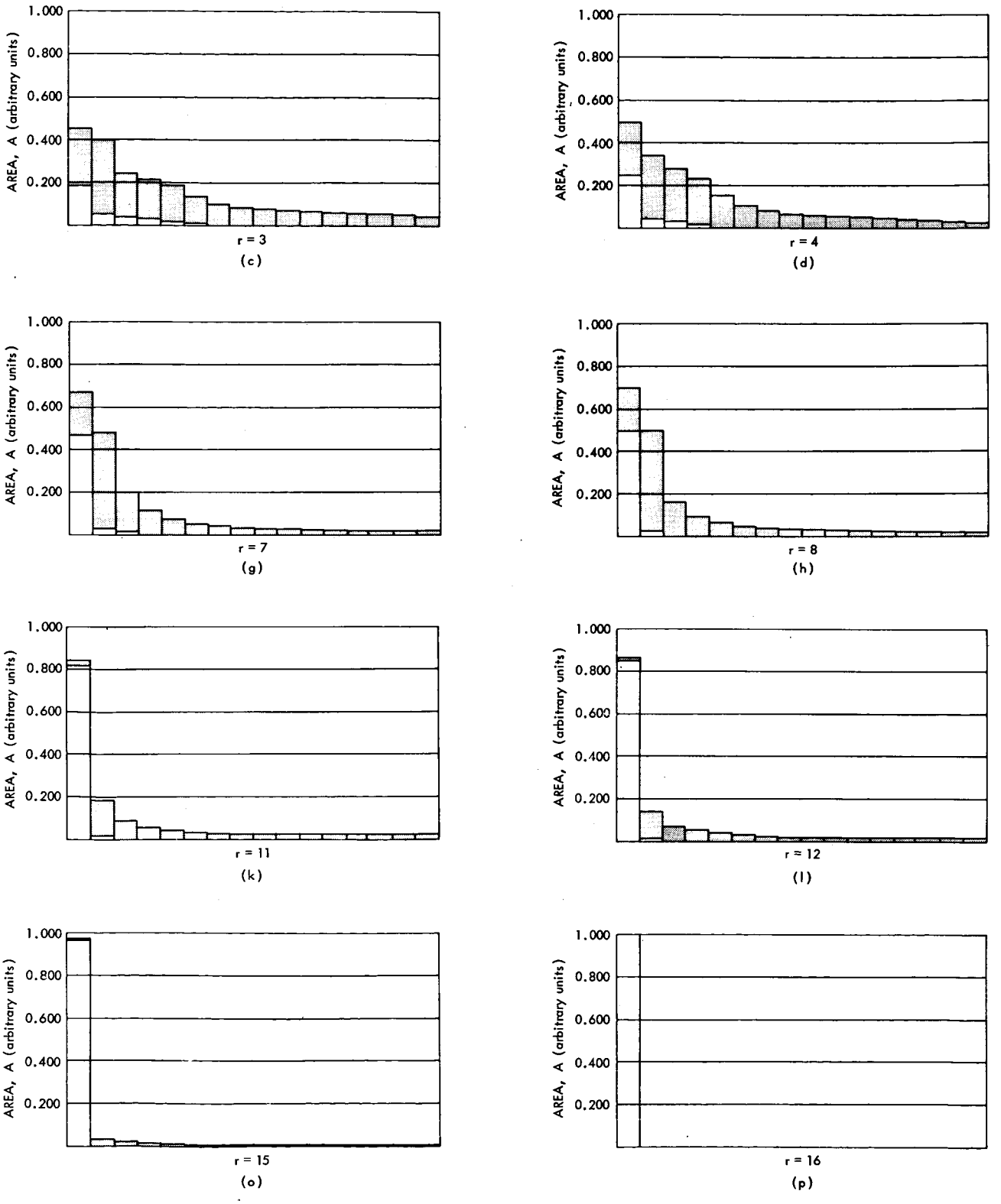


Figure 8-10(a through p)– Largest and smallest bars for histograms with the indicated ratios. (The bars of any histogram with the indicated ratio must lie in the shaded region.)

ABITS	AREA..	MIN.	MAX.	H.M.	P.E.
189 (275)		29696.00	30719.00	30198.84	1.69

SBITS	RATIO..	MIN.	MAX.	H.M.	P.E.	C(1)..	MIN.	MAX.	H.M.	P.E.
0		1.00	1.24	1.11	10.53		1856.00	5527.94	2778.97	49.73
1		1.12	2.45	1.54	37.26		2099.35	10883.28	3519.75	67.66
2		2.26	3.67	2.80	23.81		4216.92	14071.71	6489.20	53.88
3		3.40	4.89	4.01	18.00		6675.80	16582.46	9519.30	42.59
4		4.53	6.11	5.20	14.77		8991.18	18722.12	12148.24	35.11
5		5.67	7.32	6.39	12.70		11664.00	20618.53	14899.36	27.74
6		6.81	8.54	7.58	11.27		13582.24	22339.56	16893.43	24.38
7		7.95	9.76	8.76	10.22		14799.78	23926.42	18287.66	23.57
8		9.09	10.98	9.94	9.42		20319.01	25406.30	22579.65	11.13
9		10.22	12.19	11.12	8.79		22676.71	26798.30	24565.83	8.33
10		11.36	13.41	12.30	8.27		24473.25	28116.44	26168.66	6.93
11		12.50	14.63	13.48	7.85		25983.63	29371.36	27573.83	6.12
12		13.64	15.85	14.66	7.49		27312.31	30571.37	28850.09	5.63
13		14.78	16.00	15.36	3.98		28512.40	30719.00	29574.60	3.73
14		15.91	16.00	15.96	0.27		29615.28	30719.00	30157.05	1.83

ABITS	AREA..	MIN.	MAX.	H.M.	P.E.
190 (276)		30720.00	31743.00	31223.12	1.64

SBITS	RATIO..	MIN.	MAX.	H.M.	P.E.	C(1)..	MIN.	MAX.	H.M.	P.E.
0		1.00	1.15	1.07	7.19		1920.00	5009.03	2775.95	44.58
1		1.05	2.29	1.44	37.20		2030.29	10720.40	3414.01	68.15
2		2.12	3.43	2.62	23.72		4188.21	13963.10	6443.65	53.85
3		3.18	4.57	3.75	17.91		6125.34	16498.58	8933.86	45.85
4		4.25	5.71	4.87	14.67		8780.43	18652.75	11940.22	35.99
5		5.31	6.84	5.98	12.60		10219.43	20558.76	13652.44	33.59
6		6.38	7.98	7.09	11.17		13444.01	22286.61	16771.13	24.75
7		7.44	9.12	8.20	10.12		14794.25	23878.52	18269.43	23.49
8		8.51	10.26	9.30	9.32		19233.88	25362.29	21877.00	13.74
9		9.57	11.39	10.41	8.68		22174.21	26757.35	24251.14	9.37
10		10.64	12.53	11.51	8.17		24183.83	28077.93	25985.80	7.45
11		11.71	13.67	12.61	7.74		25814.02	29334.83	27462.03	6.38
12		12.77	14.81	13.71	7.39		27222.25	30536.44	28784.26	5.74
13		13.84	15.95	14.82	7.08		28480.20	31689.49	29999.26	5.33
14		14.90	16.00	15.43	3.55		29627.66	31743.00	30648.87	3.45
15		15.97	16.00	15.98	0.10		30689.47	31743.00	31207.34	1.69

has been able to increase the effective amount of information that can be transmitted to Earth from this experiment by more than an order of magnitude.

REFERENCES

1. Wilkerson, T. D., and Ogilvie, K. W., "Plasma Experiments for Space Vehicles," TN BN-378, Univ. of Md., October 1964.

2. Snively, J. W., Jr., "A Statistical Data Compression Technique," NASA TN D-3616, October 1966.

3. Schaefer, D. H., "Logarithmic Compression of Binary Numbers," Proc. IRE, 49(7): 1219, July 1961.

9. USE OF A STORED-PROGRAM COMPUTER ON A SMALL SPACECRAFT

R. A. Cliff
Goddard Space Flight Center
Greenbelt, Maryland

SMALL SCIENTIFIC SPACECRAFT

The primary purpose of a scientific spacecraft is to provide a platform from which observations and measurements can be made. The ideal scientific spacecraft would minimize the hardware requirements imposed on each experiment package. The majority of scientific spacecraft are small, weighing perhaps 150 pounds, and their attitude is spin stabilized about the axis of the highest inertial moment. Since many of the experiment sensors are directional, the spinning spacecraft causes them to perform a circular scan, which is frequently desirable. The spinning does pose a problem, however, in that the experimental data are most meaningful when collected in synchronism with the spacecraft spin, whereas the time-division-multiplex telemetry system commutates the experiments at a fixed rate that is not related to the spin rate.

Another problem is that the raw data from the many experiments that measure random events (energetic particle experiments, for instance) require a wide bandwidth but a low data content. Raw data from other experiments may not require as much bandwidth, but the average data rate may be higher. In either case, the raw data would impose a severe burden on the telemetry transmitter if all of it were to be transmitted. For this reason, devices have been included on board scientific spacecraft to accomplish buffering and simple data compression such as logarithmic counting of particle events.

Traditionally, the processing of raw data on board spacecraft has been accomplished by special-purpose equipment within each individual experiment package. Unfortunately, as the computation complexity and sophistication has increased it has become increasingly difficult for the experimenter to successfully develop an experiment system before it is obsolete. In addition, much duplication of capability occurs throughout the spacecraft. This is wasteful of power, weight, and space which a small spacecraft cannot afford.

In order to improve this situation, many advantages would result from making a small, programmable, general-purpose digital computer a part of the spacecraft electronics. The computer would do experimental data computations, data compression, and buffering. The experiment packages would then need to consist of sensors and signal conditioners only. Not only should the overall performance of the spacecraft be improved, but the time and manpower necessary to construct and check out an experiment would also be reduced.

SPACECRAFT DATA SYSTEMS

Early Interplanetary Monitoring Platform (IMP) Spacecraft

The data system as found, for instance, on the IMP series of spacecraft should be examined briefly. Figure 9-1 is an idealized view of a data system which was used on the early IMP spacecraft (IMP's A, B, and C). It should be noticed that the experiments are connected to the transmitter through a subsystem labeled "Telemetry Encoder." Data processing, if any, is performed in the experiments. (There is one exception; the telemetry encoder contains accumulators, not shown, which count pulse data from certain experiments. This is a rudimentary form of the data processing.)

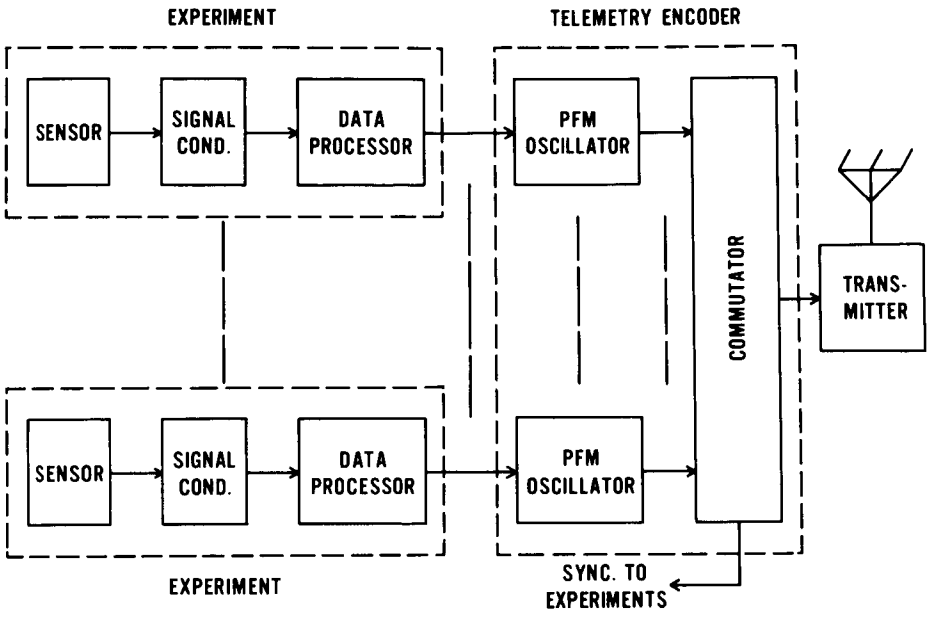


Figure 9-1—Early IMP data system

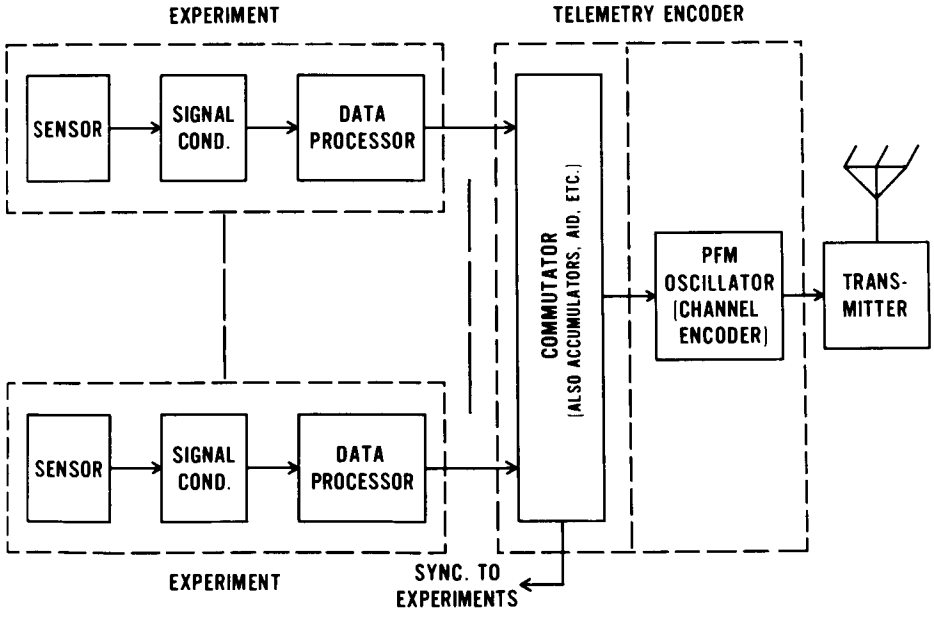


Figure 9-2—Present IMP data system

Inside the telemetry encoder are a number of PFM oscillators (Reference 1), one for each experiment. Each oscillator encodes the output data from its associated experiment as a frequency, and these frequencies are sequentially switched to the transmitter, one at a time, by the commutator. Synchronization pulses for the experiments are also supplied by the commutator.

For experiments which have analog outputs, the PFM oscillator is voltage-controlled. For experiments with digital outputs, a digitally controlled oscillator is used that produces a discrete frequency for each input state. In either case, the frequency output phase is uncontrolled and the frequency is only determined approximately. This sort of system is not optimum for a number of reasons.

Present IMP's

Performance of the system can be improved by using a phase-coherent digital PFM oscillator for all experiments. Such an oscillator is available (Reference 2) and will be used on IMP's D, E, F, and G. The new, more complete oscillator replaces the many oscillators of the past by a single unit. Data are commutated to the single oscillator which then feeds the transmitter directly.

There are disadvantages to this new method, however, for if the single oscillator fails all the data are lost, but the advantages far outweigh the disadvantages in that the new oscillator can be made more reliable than the old oscillators and fewer components are required. That the performance of the telemetry system using the new oscillator closely approaches the theoretical limits of the PFM oscillator is another advantage.

Figure 9-2 shows the data system used on present IMP spacecraft (IMP's D, E, F, and G). Except that the PFM oscillator and commutator functions have been interchanged within the telemetry encoders, it is like the data system used on the early IMP's. The PFM oscillator has been alternatively designated "channel encoder" to emphasize that, although a PFM oscillator is used, any type of channel encoder could be substituted (for instance, the pseudonoise PCM type). (The term "channel encoder" is used by communications specialists to describe a device used to encode data expeditiously in order to combat noise in the communication channel.)

Everything that the "telemetry encoder" does, exclusive of channel encoding, has been consolidated into the box called "commutator" in Figure 9-2. Besides commutation, analog-to-digital conversion (necessary now that analog voltage-controlled oscillators are no longer used), accumulation for pulse data, and the generation of synchronization pulses are included.

Future Spacecraft

The next logical step in the improvement of the spacecraft data system is shown in Figure 9-3. Again the block containing the commutator has been moved to the left. This time the multiple data processors in the various experiments have been supplanted by a single box labeled "computer." As before, improved performance is to be obtained by consolidating many similar functions scattered throughout the spacecraft into a single more efficient function. Many advantages of the centralized computer have already been discussed.

It is proposed that the basic spacecraft configuration shown in Figure 9-3 be used in future spacecraft. Certain refinements should be added, however. Figure 9-4 shows an improved spacecraft data system using a centralized computer that would be suitable for spacecraft such as the proposed Omnibus IMP's. This system has two separate encoders, each of which performs the same function as the encoder in Figure 9-3. The two encoders in Figure 9-4 are identical except for their sources of synchronization.

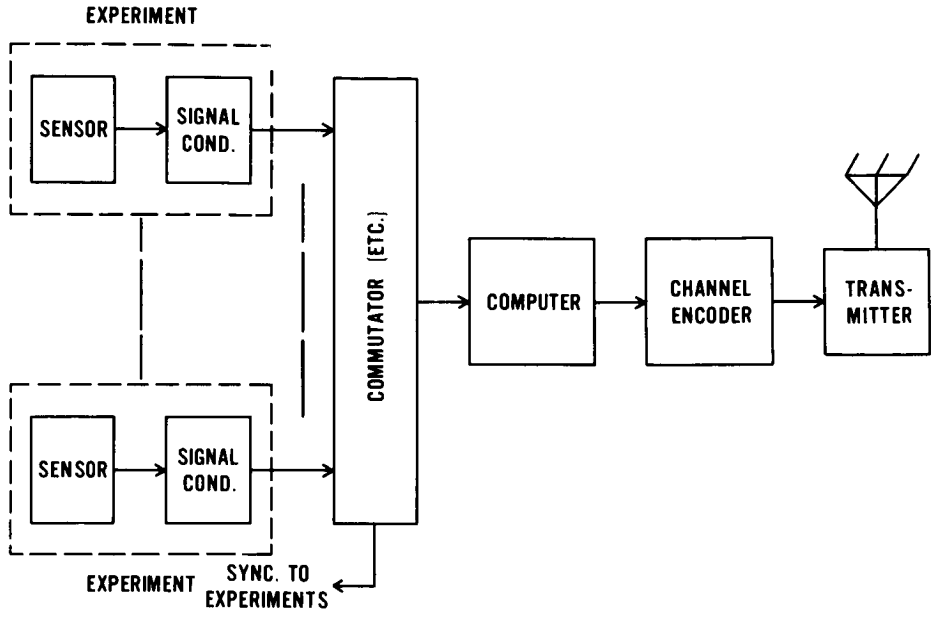


Figure 9-3—Data system with computer

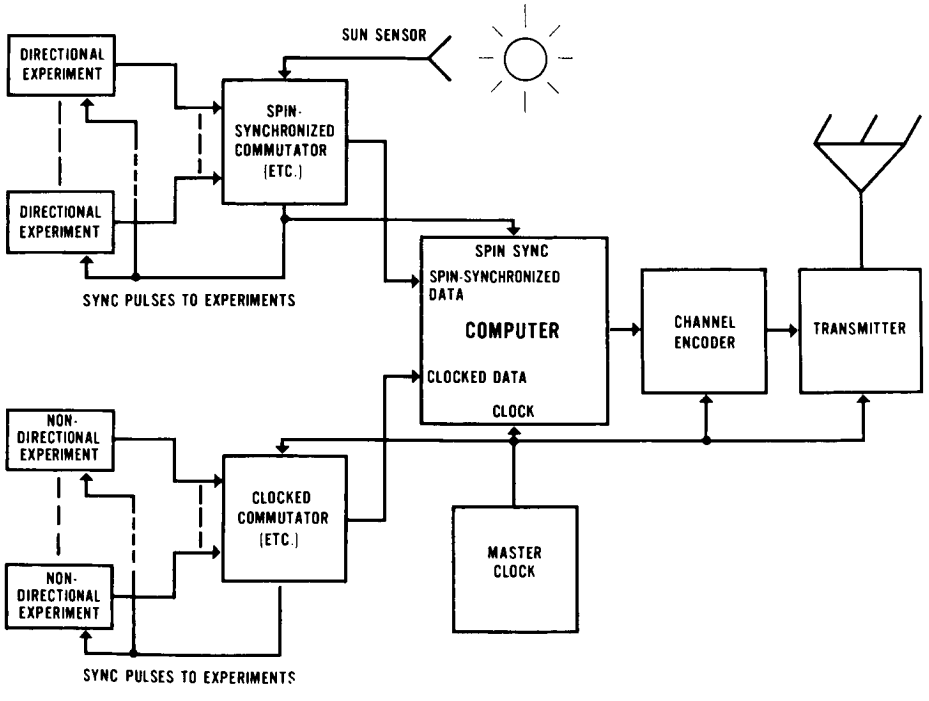


Figure 9-4—Use of spin synchronized commutator

The lower encoder (labeled "clocked encoder") operates in the conventional manner and is controlled by a fixed clock. This encoder handles the data from the experiments with nondirectional sensors. It does the usual commutation, analog-to-digital conversion, accumulation, and synchronization pulse generation. Data from the clocked encoder go to the computer for redundancy removal, data reduction and analysis, and formatting before being sent on to the channel encoder.

The significant feature which makes the spacecraft data system configuration of Figure 9-4 superior to that of Figure 9-3 is the upper encoder. It is synchronized to the spacecraft spin rate by Sun pulses supplied from the optical aspect system. For each revolution of the spacecraft, the spin-synchronized encoder goes once through its format. The advantages for experiments with directional sensors are greater because the entire operation of such an experiment is now synchronized with the spacecraft spin by pulses from the encoder, the experiment need contain no special provisions to collect certain data during particular portions of a spin. Also, data can be collected on every spin, instead of there being a collection of data during one spin and then a wait while the telemetry system reads out the data. Furthermore, the computer does the buffering between the variable spin rate and the fixed telemetry rate.

Additional possibilities exist in the ability to synchronize the operation of any (or all) of the directional experiments to any particular direction (to the Earth, Moon, or Sun, for example). It would also be possible for the spin-synchronized encoder to complete one format for any given integral number of spins. For example, if there were five directional experiments, each of which collected large amounts of data during a spin, it would be possible to allocate to each experiment a spin of its own during which it could utilize the full capability of the encoder and computer. In that case five spins would be required to complete a format.

A few more comments about Figure 9-4 are in order. If the directional stimulus (in this case the sun) that controls the spin-synchronized encoder is lost, as, for instance, by an eclipse, then the encoder should free-run at the rate it had before the loss of stimulus. Techniques for accomplishing the required synchronization characteristics have already been developed for the IMP spacecraft series (Reference 3).

Another improvement is that a master clock controls all subsystems that are not spin-synchronized. This is particularly desirable for the transmitter carrier and the channel encoder (and for the computer also) because a carrier-coherent telemetry system has synchronization characteristics superior to a noncoherent system (Reference 4). It is also important that the clocked encoder and the channel encoder be controlled by a common clock because if they are not the computer must perform an additional buffering function unnecessarily.

REFERENCES

1. Rochelle, R. W., "Pulse Frequency Modulation Telemetry," NASA Technical Report R-189, January 1964.

2. Cliff, R. A., "Digital Oscillator," NASA-GSFC Invention Disclosure, 1963.

3. Pyle, E. J., "Aspect Determination System for the Super Interplanetary Monitoring Platform," M. S. Thesis, Univ. of Md., 1966.

4. Saliga, T. V., "Using Carrier-Coherent Telemetering Rates To Improve Synchronization," (to be published).

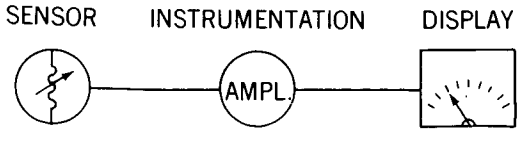
10. UNIFIED INFORMATION PROCESSING TELEMETRY SYSTEM

C. J. Creveling
Goddard Space Flight Center
Greenbelt, Maryland

Dr. Robert H. Goddard said, "It is difficult to say what is impossible, for the dreams of yesterday are the hope of today and the reality of tomorrow." The system to be described in this paper shows potentialities for approaching the "dream system" discussed by Cliff in a preceding paper of this symposium* more closely than might have been suspected possible just a short time ago. Let us first examine a few statistics**. There were 18 satellite launches in 1966 and more than 60 are planned between now and 1970, including backups. Second, it should be noted that one of the earliest satellites, the Vanguard, weighed something on the order of 10 pounds while the OGO satellite weighs over 1,000 pounds. There is a spread of two orders of magnitude here. Third, bit rates of the earliest Goddard satellites (and even some of the present ones) run as low as 20 bits per second and as high as 128,000 bits per second, or four orders of magnitude higher. The error rate for the sample taken by these satellites at the design goals initially specified runs from 10^{-2} to 10^{-8}, or six orders of magnitude. It is important that some of these facts be kept in mind when adaptive systems are discussed because no present system can cope with these ranges of variation.

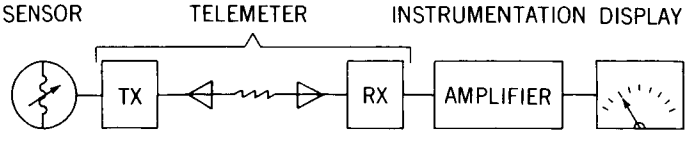
A unified information processing telemetry system will consist, basically, of a programmable computer onboard the spacecraft and an adaptive telemeter in which coding, power, bit rate, and format are controlled. It will include a programmable ground system which will have at least a capability of "talking back" to the satellite. The writer believes the term "adaptive telemetry" has its widest meaning in this sense. A multidisciplinary approach must be considered in order to achieve any solutions to the problems which cover such a gamut. One cannot take the position of a telemetry engineer who considers his system as a common carrier and deals with its inputs and outputs according to some specification.

Figure 10-1 depicts a system for a simple laboratory experiment that requires one system engineer – the experimenter. Figure 10-2 shows what happens when the output of a sensor is telemetered from some distance away because of safety considerations or the impossibility of having the experimenter present. Here a second man comes into the picture – the subsystem engineer.



SYSTEM ENGINEER: EXPERIMENTER

Figure 10-1—An experiment in the laboratory



SYSTEM ENGINEER: EXPERIMENTER
SUB SYSTEM ENG. : TELEMETRY INSTRUMENTATION

Figure 10-2—An experiment on the range

*Paper 9, pp. 129 to 133.

**From "Major Space Projects at Goddard," internal GSFC document (Official Use Only).

Figure 10-3 shows an increasing order of system complexity that requires that an increasing number of people become involved in the design. Now the problem can be viewed as a whole and information systems can be discussed. Although data handling inside the subsystems is of concern, the meaning of the data as it passes through the system as a whole is of equal concern. This best illustrates what the term "information" means when an information system is discussed. The overall system has a number of subsystems and it would be very difficult to present in detail adequate specifications of the inputs, outputs, and interfaces between subsystems. Until each subsystem engineer concerns himself with the subsystems which adjoin his, and indeed with the whole satellite plan, adequate control will be lacking. For example, consider an experimenter looking at the pulsed output of a sensor in a simple experiment. Initially, the experimenter would like to see not only the pulse but the wiggles on the pulse to establish in his mind the integrity of the instrumentation. Only after he has been convinced that his experiment is indeed working properly is he interested in the pulse value. If the system gives him this value and nothing else, he will still question the integrity of the experiment. If the experimenter is asked what he would like to have in a telemetry system, he might say, "Give me 5 megacycles!" It is believed that it is possible, with the use of onboard processing and an adaptive system, to satisfy both of these requirements in the sense that they can be controlled from the ground. To do this an initial learning period is provided in which any or all of the experiments can be sampled successively at very high rates in order to establish confidence in the system, and then only the desired information is transmitted.

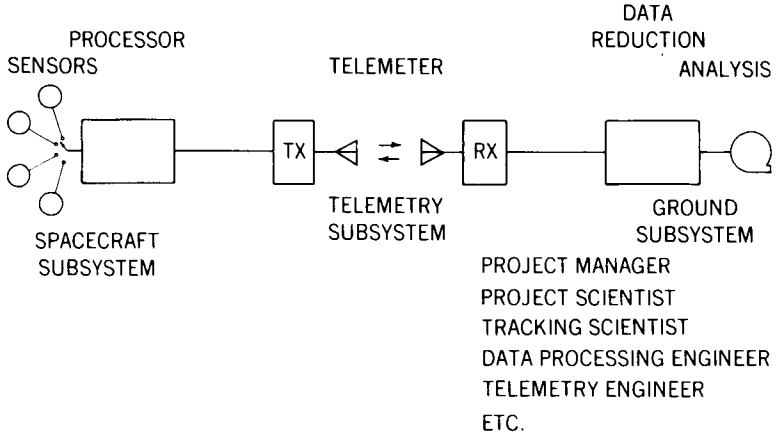


Figure 10-3—Subsystems used in coupled experiment

A great deal has been said about the problems of coding as a technique used in various communications systems. Many of these codes, however, have been generated to satisfy specific problems but our concern here is with the error rates covering six orders of magnitude. It is very difficult to imagine any single set of codes that would satisfy the requirements over this range; however, once there is a computer on board the spacecraft and we have the means for controlling it, the computer becomes a very useful adjunct to coding the telemetry link. Therefore, the computer belongs as much to the telemetry engineer as it does to the onboard processing system.

A system is being proposed by the GSFC Information Processing Division over which they have to exercise a measure of unified control in its design. Unified control means that all of the subsystems will be under a single administration, which will provide a fairly close control over both the administrative and technical problems involved in developing all the various subsystems, from the spacecraft to the ground support. However, when it is necessary to become involved in an interdisciplinary approach, the semantic problem immediately arises of communication between people who at times use the same terms with different meanings and who use unfamiliar terms. It is difficult, for example, for a person who is not used to working with computers to understand what

goes on inside the computer. A thorough understanding of a computer program can be obtained neither by looking at the punch cards nor by looking at the sequence of the machine instruction codes. In fact, it is doubtful that a complete understanding can be obtained by reading the program in some language such as Fortran. To get this understanding, not only must the program be expressed in some relatively higher order language but also an English description of it and a flow diagram are necessary. A number of so-called higher order languages have been developed for computer programmers in place of, or to supplement, flow charts and the English descriptions which can describe very complicated systems very precisely. For example, the logic structure of the IBM system 360 has been completely specified in a higher order language in approximately a dozen pages. The programming that takes place on the IMP-F satellite, which is a relatively small system, was described by Dr. E. P. Stabler* in this higher order language in a few sentences.

It is presently proposed to study and use such a higher order language for several reasons. First, it is most convenient for conveying information from one subsystem designer to another in a complete and unmistakeable manner; second, this language lends itself to the writing of the actual programs which will be used in the computer on board the satellite and in the computers on the ground (although they are different types); and, third, this language becomes a design tool in the design of the onboard computer because it is possible to write the Boolean expressions for the functions expressed in higher order language (in fact, this has already been started). From these expressions the logic diagrams can be made through a relatively straightforward process.

It is not clear just how useful this language will be to people outside the information system. It would be naive to think that, because this wonderful thing has been developed and discovered, everybody will say, "Teach it to me." What system engineers have to do first is use it for their own benefit. However, these languages have been in vogue and widely distributed now for several years and are finding some use by programmers in general. Moreover, almost all satellite experimenters have been forced to become computer-oriented. This provokes the increasing use of these languages as a natural trend.

The kind of machine that is being proposed here for spacecraft computers follows work that was done by Dr. E. P. Stabler at GSFC last summer. At that time it was undertaken to design a stored program computer that would perform the same functions as one of the IMP spacecraft processors, and it was concluded that one could be designed with approximately the same size, weight, and power consumption that would be busy less than 10 percent of the time. Therefore, it should be possible to build a computer for small satellites that would have a considerable amount of time left for things such as coding the telemetry link, changing the format of the data passing through the system, and possibly processing some of the information before it goes into the telemetry link. If there is a computer on board the satellite, a computer on the ground, and a command system, a "dream system" starts to take shape – a dream system in which a spacecraft ultimately might be considered a piece of peripheral gear used in conjunction with a large ground-based computing system. This spacecraft computer would then perform experiment conditioning, multiplex processing, and coding for the telemetry link. In addition, the modularity in design necessary to overcome problems in reliability would enable the configuration to be changed from one mission to another.

Figure 10-4 shows in a very simple manner a microprogrammed machine and a stored program computer. These differ as shown in Table 10-1. The conventional computer is characterized by fixed input format and fixed word length. On the other hand, in the microprogrammed machine variable word length can be provided very inexpensively. The conventional computer has a fixed logic structure, a stored program, and a fixed instruction repertoire that can be set up on the ground. The microprogrammed computer differs in this respect in that the microprograms as well as the macroprograms are stored and are changeable. In the conventional computer, flexibility is provided by software and the capability for simultaneous operations is limited. In general,

*Associate Professor of Electrical Engineering, Syracuse University.

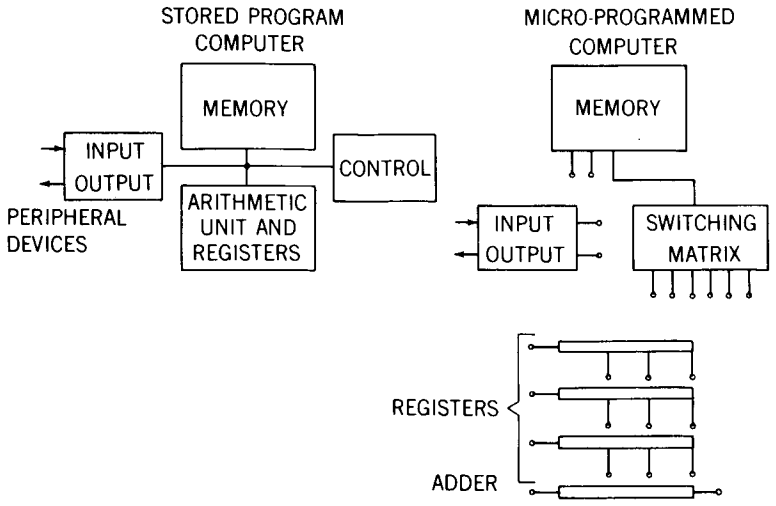


Figure 10-4—Computer configuration

Table 10-1 – Computer Characteristics

Conventional Computer	Microprogrammed Computer
Fixed input format; fixed word length	Variable input format
Fixed logic structure; stored program and fixed instruction repertoir masking for bit manipulation	Flexible internal structure
Large stored program memory	Small stored program
Low speed (many instructions per macro)	High speed
Flexibility provided by software	
Limited simultaneity	

a large stored program memory or else a very extensive list of instructions is required and it is also low-speed in the sense that many machine instructions have to be carried out for each micro-program or so-called macroinstruction. The microprogrammed computer, however, circumvents many of these limitations by being, in effect, almost any machine that you want it to be, and its instructions, in effect, rewire the machine for the problem at hand. The number of instructions to be stored in order to do this is surprisingly modest and the speed is very good.

Development of this system is now progressing in a series of steps. The study phase is in process and a design phase has been started in which the tentative logic designs are being carried out. A breadboard is planned for testing the actual onboard subsystem. It is not necessary to build a computer on the ground because there are a variety of these available and they can be programmed according to higher order language. In order to see how this system will operate as a whole, a computer simulation of the whole system has been planned. This is not going to be an easy job but it can be accomplished in parallel with the development and later on it will be invaluable in reconfiguring the system according to what is learned from the simulation. Having a flexible onboard computer permits the feedback of simulation results and early experiences with the system. This will be followed by the development of a prototype and a flight model in the future.

SUPPLEMENT QUESTION AND ANSWER SESSION

UNIDENTIFIER QUESTIONER: I have two things, one is in the nature of a comment. Although the computer is a very useful tool, I think you have to be careful not to use it for things that it is not emminently suited for. For instance, you mentioned perhaps doing a channel coding in the computer, which in certain instances could be a good idea and certainly provides lots of flexibility. However, I think this is one thing that is quite efficiently done by fairly elegant hard-wired devices which have been developed such as the PN system. To go another step further, if you are counting data pulses from an experiment-interrupt computer and have to add "one" to a register, you could include special accumulators to do this and then transfer the results entirely. You have to be somewhat careful what jobs you do assign to a computer.

The other thing I have is a question about the microprogrammed machine. Do you have any feeling for the difference in complexity between the standard configuration and microprogrammed configuration, especially considering this is a box-labeled switching matrix and may grow to fairly considerable proportion?

MR. CREVELING: I think your remarks are certainly well taken and reflect my feeling on the subject of the software approach versus the hardware approach to specific problems within the spacecraft, such as coding the telemetry link. I mentioned the fact that the computer is available and that the computer is a very flexible device so that as you change from mission to mission, or if you change from very close ranges to very distant ranges as in some of our eccentric satellites, you have the ability then to change your type of coding. It is certainly true that if you have a fixed coding scheme you could build up hardware devices to do this very efficiently and very quickly. In such cases, we have generally followed the practice of investigating both approaches, trying it both ways (at least in the study phase) to decide which is the better, and making the choice on that basis, taking into account the fact that sometimes expediency calls your hand. I feel that the reason that the computer has advantages for doing this is not because of greater efficiency but because of the flexibility of the approach.

Regarding your question as to the difference between the microprogrammed machine and the more conventional machine, we plan to try both of those, and compare them. At the moment I could not give you a comparison because we have not gotten that far in our study, but I think that the microprogram machine will compare favorably with the conventional approach.

MR HABIB: In your discussion here a thought occurred to me. I wonder whether we did not miss the point (semantics again) dealing with the word "computer" when we are really talking about general-purpose data or signal-handling system. You began your talk by talking about unified information systems and the designer needed in a sense a control over all of the elements in it. Then we very suddenly get down to a specific thing and call it a computer and I think what flashes into everybody's mind is the type of computer that sits on our floors out here. I think you both were correct that you will use unique equipment words where needed and use general-purpose words where needed and so long as you are allowed to design the entire system then you will achieve your goals.

MR CREVELING: No argument.

DR. POSNER: If you were to use your computer to encode the telemetry, I envision that a computer failure in the central computer clock would cut off all information as to what happened and, in fact, by having all experiments channeled through one machine you may find out that it becomes a no-go instead of an OGO spacecraft. In this case you may be very unhappy at having everything channeled through one point. We at JPL feel that decentralization will protect a mission. Now our missions may be more expensive and you may have to count on getting some use out of them. On a cheaper mission it may be that the reliability is not as important. These are sort of political decisions but in the very expensive heavily equipped spacecraft for soft landing on Mars you have to count on getting some experiments back no matter what fails, so, in a really

expensive mission where you have everything staked on it, I would hestiate to think of channeling everything through one computer. On a smaller interplanetary spacecraft where you have dozens of them in the warehouse you may find it more economical to use the central computer. Incidentally, a possible use for a computer once you have it is to decode the commands from the ground. You may want to have very highly encoded commands to avoid having a miscommand decoded by the spacecraft which might 'wipe out' the mission. If you do have a general purpose computer, that would be a really good use for it. Perhaps using a computer to decode spacecraft commands may be a better thing to do than use it to encode telemetry.

MR. CREVELING: That is one of those bad little nightmares that sometimes interrupts our dream system and we have thought of it. If you will notice in Mr. Purcell's diagram, he shows a switch by which his airborne computing system could, in effect, bypass the data from the commutator (which presumably would remain unimpaired) and would continue to send some kind of data down to the ground. There is another approach, several varieties of approach, to this reliability problem. They generally involve the use of some kind of equipment redundancy. One form which has been used in a large-scale computer is the one in which you have a number of memory cells and a number of computing units, and these two are connected by a matrix. If one or more of these fails, you limp along with something less than the capability of the original system. Since it is programmed, you now reprogram to handle less information. I hope that we can use some such technique to remove the fears that you talk about and, as to your suggestion on using the computer to decode, it is a very good idea.

11. CHANNEL NOISE – A LIMITING FACTOR ON THE PERFORMANCE OF A CLASS OF ADAPTIVE TECHNIQUES

G. L. Raga

Electro-Mechanical Research, Inc.
Sarasoto, Florida

INTRODUCTION

The present paper reports the results of an experimental program devoted to the determination of the effects of channel noise on a class of adaptive techniques. Solutions to certain problem areas are postulated and the net efficiency of the coding techniques is determined.

The field of adaptive-compression telemetry techniques is relatively new compared with areas such as television and speech-bandwidth compression. One of the earlier papers on the subject, whose central theme was the design of an efficient telemetry system, appeared in 1959 (Reference 1). Since that time, effort has been concentrated in four major areas: Adaptive sampling, preprocessing, selective monitoring, and efficient encoding.

From the available literature it appears that the majority of the users and investigators of telemetry compressions are considering the zero-order interpolator. An unfortunate discrepancy exists in the terminology in that the zero-order interpolator is frequently described as a zero-order predictor (Reference 2), self-adaptive compression (Reference 3), an adaptive sampling technique (Reference 4), selective monitoring (Reference 5), run-length encoding (Reference 6), floating aperture (Reference 7), redundancy removal (Reference 8), and the step method (Reference 9).

These techniques, although sounding basically different, operate identically. Although they have been applied to a telemetry source, they can also be applied to a television source. For the purposes of illustration, the source will be considered a television source. Start with word 1, line 1, of a television frame. Word 1 is compared with word 2; if the absolute value of the difference is less than K, word 2 is disregarded and the process continues until the difference is greater than K, say at word j. At this point, word 1 is transmitted to a buffer store along with the distance from word 1 to word j. Word j then replaces word 1 and is used as a basis, and the process continues. It is this class of adaptive techniques that will be investigated.

PROBLEM DEFINITION

Adaptive techniques for data transmission can be applied to various sources, including video, speech, and telemetry data. There are three possible ways to take advantage of the adaptive technique: (1) if the bandwidth is fixed, adaptive-compression techniques can save transmitter power; (2) if the transmitter power is fixed, bandwidth can be conserved; and (3) if both the transmitter power and bandwidth are fixed, more information can be transmitted per unit time. Assume, for example, that photographs are transmitted from an earth-orbiting vehicle with rf parameters as given in Table 11-1. With the channel capacity fixed at $A = 10^6$ bits per second and by employing data compression, more information

Table 11-1

COMMUNICATION-LINK PERFORMANCE

Parameter	Symbol	Value	Value
Transmitter power	P_t	2 watts	3 db
Transmitting antenna gain	G_t	Omni	0 db
Receiving antenna gain	G_r	60' Parabolic	49.9 db
Receiving system temperature	T_r	200° K	
Transmitting frequency	f	2300 MHz	
Receiving noise spectral density	ϕ_k		-205.6 dbw/cps
Range	R	5000 n mi	
Free space loss	L_s		179.0 db
Miscellaneous spacecraft loss	L_m		3 db
Design margin	L_t		6 db
Bit rate	R_b	10^6 bps	60 db
Energy/bit to noise spectral density	E/(N/B)		10.5 db
FM bandwidth	B_{if}	700 kHz	58.5 db
Carrier-to-noise ratio	C/N		12 db

can be received per unit time. If the PCM television frame is composed of x samples per line and y lines per frame, each sample is quantized to n bits per sample, and F frames per second can be transmitted, the data rate is

$$R = Fnxy \text{ bits/second.} \tag{1}$$

The number of data frames that can be transmitted is

$$F = \frac{A}{nxy} \text{ frames.} \tag{2}$$

The amount of information that can be transmitted in t seconds is $I = Ft$.

If F can be increased, I increases, and a savings results. One way of increasing F is to decrease nxy by some form of data compression.

ADAPTIVE SAMPLING TECHNIQUES

The adaptive sampling technique that will be discussed is the zero-order predictor. Assume that the frame of data is composed of xy samples. If the coding algorithm is applied, the result is a sample reduction of

$$C_s = \frac{xy}{S}, \tag{3}$$

where S is the number of nonredundant samples per frame. In order to reconstruct the data at the ground stations, it is necessary to address each nonredundant sample. By assuming that the absolute magnitude and address of each sample in a line is transmitted, the gross average data compression is

$$C_g = \frac{nxy}{\left[\log_2(x) + n\,S\right]}. \tag{4}$$

For example, consider $n = 6$ bits and $x = y = 256$. Then

$$C_g = 0.429\, C_s. \tag{5}$$

Hence, addressing the nonredundant samples results in approximately a 60 percent reduction in compression.

For the purposes of this analysis, it will be assumed that the incremental distance Δx between nonredundant samples will be transmitted, instead of the absolute address of the nonredundant samples; that is,

$$C_g = \left[\frac{n}{\log_2(\Delta x) + n}\right] C_s. \tag{6}$$

For $n = 6$ and $\Delta x = 16$, $C_g = 0.6\ C_s$.

With this form of addressing, the sample reduction compression is only 40 percent rather than the 60 percent achieved with absolute addressing.

CHANNEL NOISE

Channel noise is another limiting factor on the efficiency of an adaptive sampling technique. Under the constraints of fixed transmitter power and bandwidth, the effects of channel noise on the zero-order predictor will be more pronounced than on PCM data transmission. Each transmitted segment now represents, on the average, C_s elements. Therefore, a single bit error will affect many elements. It will be shown later that, depending upon the system implementation, a single error can destroy an entire frame of data, a single line of data, or just a few samples.

For the zero-order predictor analyzed, it will be assumed that perfect line synchronization is available and that the first element of each line transmitted is correct. With these constraints, the maximum degradation that can be caused by a single bit error is the loss of an entire line of data; however, the error cannot propagate into the next line. The most logical solution to the problem is to increase the signal power (Reference 10); however, this is not permissible under the constraints of the stated problem. Therefore, once the channel characteristics are specified, it is necessary to add error protection to minimize error propagation either by means of error detection and correction or by error detection and retransmission. For the purposes of this paper it will be assumed that the former case will be employed, that the channel is binary symmetric, and that the type of error protection will be a single-error-correcting coding of the Hamming class. In this case, the net data compression is reduced further by the amount of redundancy the error-correcting code adds. In general, if P bits is the amount of redundancy added to each transmitted word, the net data compression is given by

$$C_n = \frac{n x y}{\left[\log_2(\Delta x) + n + P\right] S} \tag{7}$$

or

$$C_n = \left[\frac{\log_2(\Delta x) + n}{\log_2(\Delta x) + n + P}\right] C_g. \quad (8)$$

which is merely a restatement of the fact that the net compression is equal to the gross average data compression times the efficiency of the error-correcting code. Carrying through the previous example, but adding a (14 and 10) Hamming code,

$$C_n = 0.429\ C_s; \quad (9)$$

that is, when error protection and addressing are employed on the system, the net data compression is approximately 43 percent of the sample reduction. With the basic terms defined, it is now possible to compare the performance of the adaptive sampling system.

The basis of comparison of the system's performance with that of other systems will be a subjective evaluation; however, so that the data can possibly be extrapolated to other sources, the rms error will also be used as a basis. In general, for the zero-order predictor the peak error is normally specified as being an error criterion; however, because of the expansion of errors due to channel noise, this cannot be used as a basis of comparison. Since this error-expansion factor is quite pronounced for the lower error rates, there is a need for another error criterion. As a consequence, the experimental program was initiated to investigate the effects of channel noise.

EXPERIMENTAL PROGRAM

The objectives of the experimental program were to evaluate subjectively the effects of channel noise on PCM and compressed data, to determine experimentally the rms error as a function of compression, and to determine experimentally the rms error between the compressed data and PCM data at varying channel bit error probabilities.

Figure 11-1 is a block diagram of the conceptual experiment. A photograph was scanned by EDITS (Electro-Mechanical Research's (EMR's) experimental digital television system). The digital data then were transferred to EMR's ASI-210 digital computer and processed by the computer with various coding algorithms, the channel noise was injected into the compressed data, and the data then were decoded and transmitted back to EDITS to be displayed and

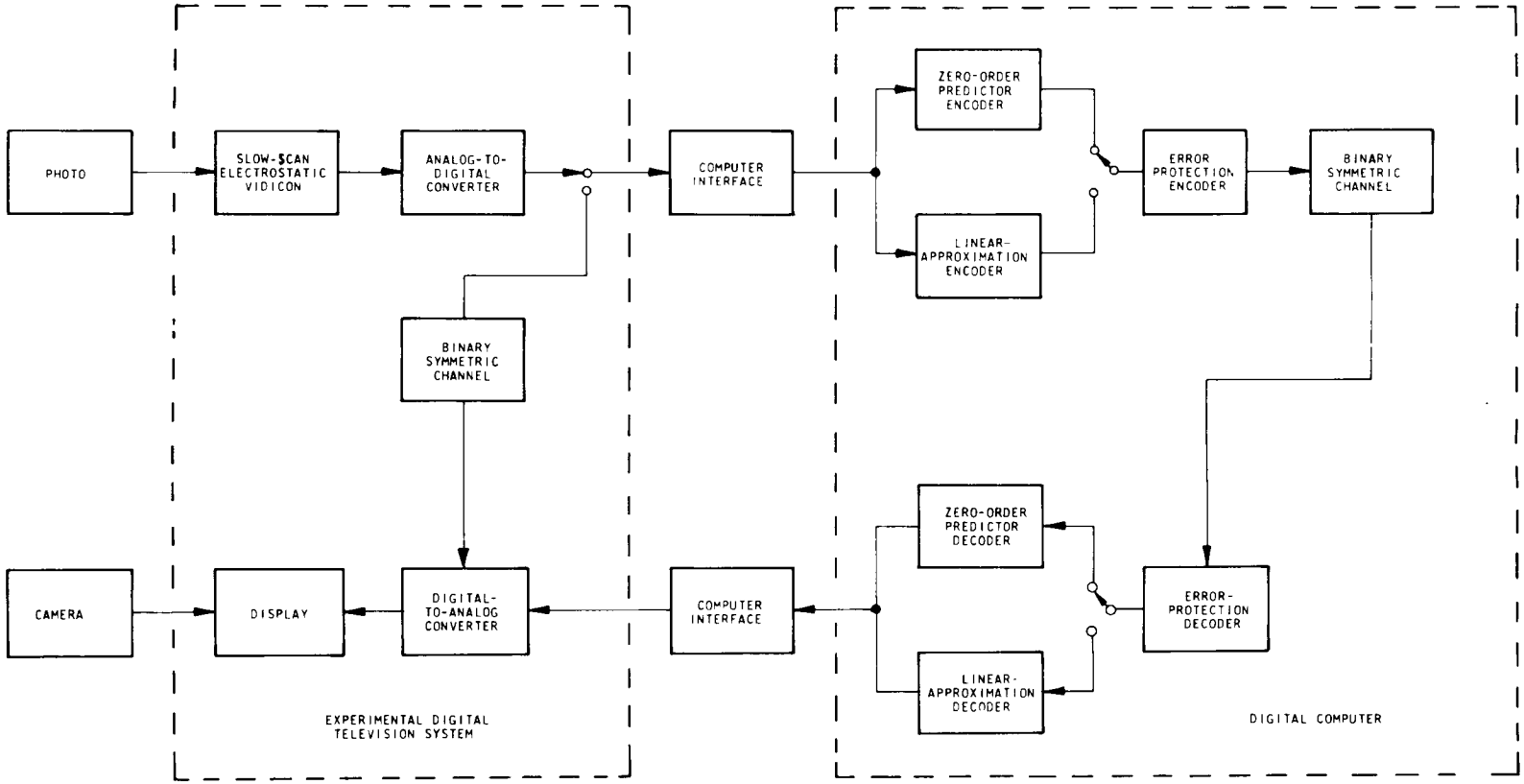


Figure 11-1. Block Diagram of Experimental Setup

photographed. A third path existed in the computer: the addition of error correction in the form of a (7,4) or (14, 10) Hamming code. The (7, 4) code, which was the easiest to implement and the fastest operationally, was used to protect only the Δx positional bits since an error in position is weighted far heavier than an error in intensity. A (14, 10) Hamming code which protected both the intensity and the positional information was also implemented on the computer. This code has the same net coding efficiency as the (7, 4) code; however, its figure of merit (the ratio of the word error probability after coding to the word error probability before coding) is not as large as that of the (7, 4) code and is an order of magnitude greater in complexity in both implementation and encoding time. However, as anticipated, the code performed more efficiently than the (7, 4) code.

ERROR PERFORMANCE

In the past, most investigators have ignored the effects of channel noise on PCM data or have argued that with compression the bandwidth is reduced and therefore the noise also is reduced. It is obvious that an error in the most significant bit has more of an effect than an error in the least significant bit and, of course, all errors are equally weighted. A derivation of the rms error as a function of the channel bit error probability is a straightforward matter. The rms error as a function of the channel bit error probability ($P_e{}^b$) and the number of quantizing bits N is given by

$$\mathrm{rms}_{PCM} = \sqrt{\frac{P_e{}^b}{3}\left(1 - \frac{1}{4^N}\right)}. \tag{10}$$

Errors in a video PCM scene occur as a salt-and-pepper effect and are constrained to individual samples; however, for the zero-order predictor this is not the case. For the zero-order predictor a single-bit error can propagate the average sample reduction, it can propagate the entire length of a line and destroy all the samples, or it can merely cause a finite percentage of the samples in the line to be in error. Assume that Δx is encoded to M bits and the intensity, to N bits; therefore, for each nonredundant sample, an $(N + M)$-bit word is transmitted.

Figure 11-2 shows the error-analysis model. State A occurs when the decoder is in perfect synchronization with the encoder; that is, when transmission is error free. State B occurs when a single bit error occurs in intensity bits; this will cause an error to propagate the average sample reduction and then return to state A. State C represents the system being completely out of

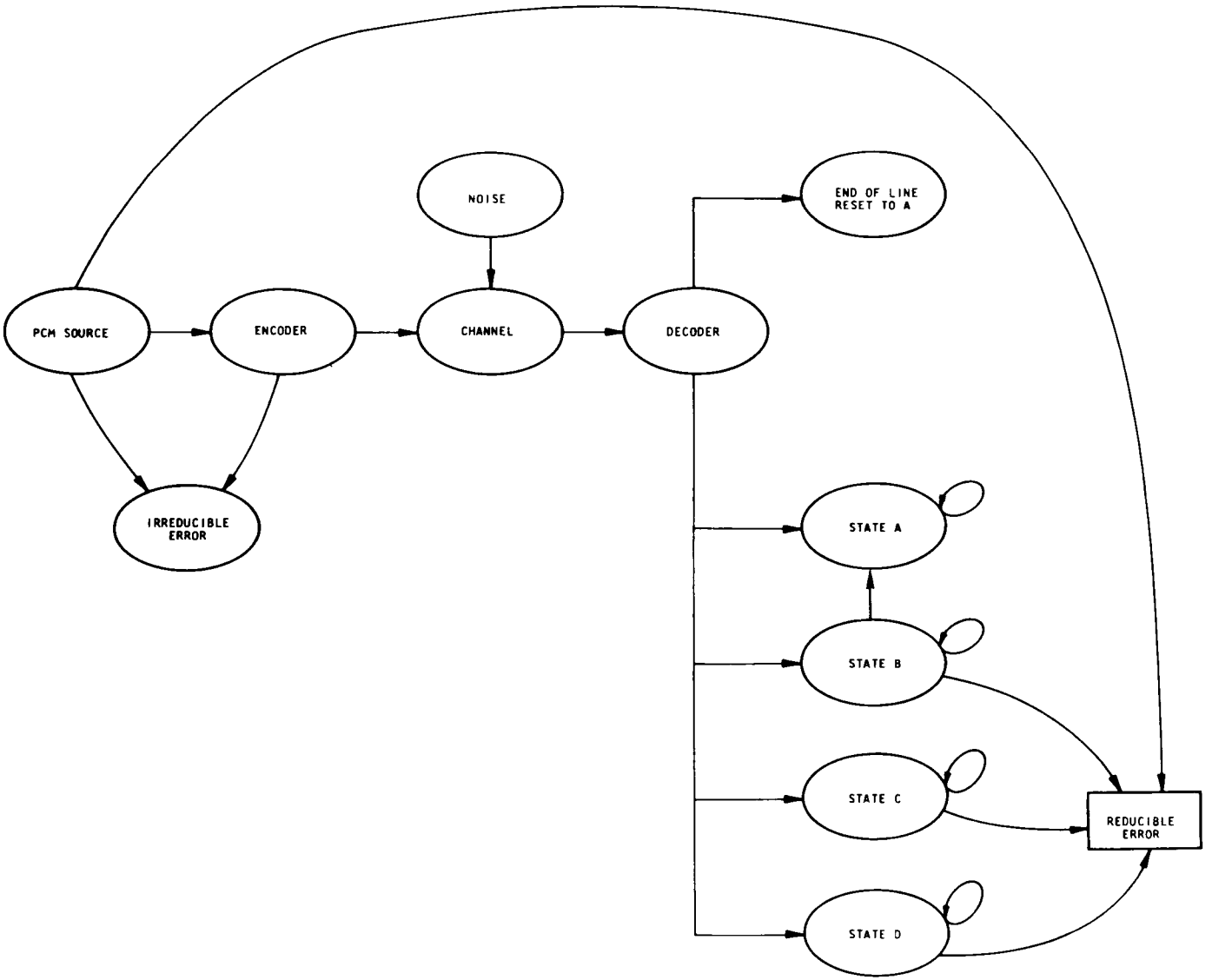


Figure 11-2. Analysis Model

synchronization; however, the error propagation is finite since this can be described as a sliding synchronization state; that is, an error in the least significant bit of Δx will cause a displacement of ±1 sample for each segment, but the remaining elements will still maintain the correct intensity since there exists an average number of successive samples of the same intensity equal to the sample reduction. Once the system enters state C, it cannot reenter state A without additional constraints on the decoder. For the purposes of this analysis the decoder returns to state A at the start of each line. State D occurs when an error occurs in the most significant bits. Such an error will cause complete loss of synchronization with zero probability of returning to state A, and, in general, a complete loss of each line occurs. As with state C, state D will return to state A at the beginning of each line by external control.

Assuming that the channel bit error probability is given by $P_e{}^b$, the probability of being in state A is $(1 - P_e{}^b)$ and the rms = 0. The probability, to a first-order approximation, of being in state B is given by

$$\text{Prob(state B)} = \frac{N}{N + M} . \tag{11}$$

The probability of being in state C or D is the same and is given by

$$\text{Prob(state C)} = \text{Prob(state D)} = \frac{1}{2}\left(\frac{N}{N + M}\right). \tag{12}$$

The rms error can be considered as an rms error-expansion factor. Assuming that the encoded frame is used as a basis of comparison, the rms error is given by

$$\text{rms}_{\text{ZOPE}} = \sqrt{\frac{1}{4x}\ (MS_A + MS_B + MS_C + MS_D)}. \tag{13}$$

Hence, an error in intensity will have an rms value of

$$\text{rms} = \sqrt{\frac{P_e{}^b}{3}\left(1 - \frac{1}{4^N}\right)}, \tag{14}$$

and it will occur (N/N + M) percent of the time and will propagate C_s samples. Thus,

$$R_{MS_B} = C_g \sqrt{\frac{P_e{}^b}{3} \left(1 - \frac{1}{4^N}\right)}. \tag{15}$$

An error in Δx that will cause the system to enter state C will occur $M/\left[2(N + M)\right]$ percent of the time and will cause an rms error

$$\sqrt{\frac{P_e{}^b}{3} \left(1 - \frac{1}{4^N}\right)}, \tag{16}$$

to be weighted by $E(S_L)$ samples/line times $E(D)$, the expected value of the displacement. It can be shown that

$$E(S_L) = \sqrt{\frac{1}{3}\left[\frac{Nx}{(N + M)C_g}\right]^2 + \frac{Nx}{2(N + M)C_g} + \frac{1}{6}} \tag{17}$$

and

$$E(D) = \sqrt{\frac{1}{2^M}\left(\frac{16^{M/2} - 1}{3}\right)} \tag{18}$$

for M, an even integer.

A single bit error in Δx that will place the system in state C will have an rms value of

$$rms_C = \frac{M}{2(M + N)} \sqrt{\frac{P_e{}^b}{3}\left(1 - \frac{1}{4^N}\right)\left[\frac{1}{2^M}\ \frac{(16^{M/2} - 1)}{3}\right]\left\{\frac{1}{3}\left[\frac{Nx}{(N+M)C_g}\right]^2 + \frac{Nx}{2(N+M)C_g} + \frac{1}{6}\right\}} \tag{19}$$

An error of Δx that will place the system in state D will have an rms value of

$$\sqrt{\frac{P_e{}^b}{3}\left(1 - \frac{1}{4^N}\right)} \tag{20}$$

occurring $M/[2(M+N)]$ percent of the time weighted by a factor $E(\zeta)$.

It can be shown that this is given by

$$E(\zeta) = \sqrt{\left[\frac{(N+M)C_g}{N\sqrt{6}}\right]^2 - \frac{(N+M)C_g x}{2N} + \frac{x^2}{3}}. \quad (21)$$

Therefore, the rms error introduced when the system is in state D is given by

$$rms_D = \frac{M}{2(M+N)}\sqrt{\frac{P_e{}^b}{3}\left(1 - \frac{1}{4^N}\right)\left[\left(\frac{(N+M)C_g}{N\sqrt{6}}\right)^2 - \frac{(N+M)C_g x}{2N} + \frac{x^2}{3}\right]}. \quad (22)$$

The total rms error expansion for the zero-order predictor is given by

$$rms_{ZOPE} = \sqrt{\frac{MS_A + MS_B + MS_D + MS_C}{4x}}. \quad (23)$$

This is only part of the total error which is caused by channel noise, and it is relative to the encoded frame. The total error is given by this error plus the rms error due to the encoding process.

RMS ERROR DUE TO ENCODING

The coding algorithm concept introduces an rms error that, because of the coding technique itself, is irreducible. For the zero-order predictor with an error band of ±K elements, in a single frame of data there are, on the average,

$$S = \frac{Nxy}{(N+M)C_g} \quad (24)$$

samples which by definition are correct. Hence, the number of possible elements that could be in error is given by

$$xy\left(1 - \frac{N}{(N + M)C_g}\right), \tag{25}$$

or $1 - \left[N/(N + M)C_g\right]$ percent of the elements may be in error. The expected value of the error is given by

$$E(\epsilon) = \sqrt{\frac{2K^2 + 3K + 1}{6}}\ . \tag{26}$$

However, of these elements that may be in error, the expected number that could be in error is determined by the amount of redundancy in the frame of data. Let PE be defined as the amount of redundancy where, as the error band increases, the amount of redundancy increases also; therefore, the percentage of elements that may be in error is given by

$$\left[1 - \frac{N}{(N + M)C_g}\right]\left[1 - PE(N)\right]. \tag{27}$$

Consequently, the rms error due to the encoding algorithm is given by

$$rms_{ZOP} = \frac{1}{2^N}\sqrt{\left[1 - \frac{N}{(N + M)C_g}\right]\left[\frac{2K^2 + 3K + 1}{6}\right]\left[1 - PE(N)\right]}. \tag{28}$$

The total rms error as a function of the channel bit error probability is given by

$$rms = rms_{ZOP} + rms_{ZOPE}. \tag{29}$$

These theoretical predictions will now be compared with actual experimental results.

EXPERIMENTAL RESULTS

Because of the nature of the source data, both subjective and rms experimental results will be given.

Subjective Results

Figure 11-3 shows a comparison of the subjective results of a zero-order predictor, the 6-bit PCM original, and the linear-approximation coding technique. For rather large apertures (error bands), contouring results for the zero-order predictor; however, the contouring can be minimized somewhat by the linear-approximation technique. The zero-order predictor can be thought

Linear Approximator
$C_s = 7.53$
RMS = 3.53%

6 Bit PCM

Zero-Order Predictor
$C_s = 6.05$
RMS = 3.54%

Figure 11-3. Comparison of Zero-Order Predictor and Linear Approximation Coding Techniques

of as a run-length (zero-slope) encoder with an error band. The linear approximation allows a finite number of slopes with an error band; therefore, it should eliminate the contouring effect as the aperture is gradually increased. Figure 11-4 illustrates the subjective effect of varying the compression for the zero-order predictor.

The effects of channel noise on a zero-order predictor are given in Figure 11-5. Here the error band is fixed and the channel bit error probability is varied from 10^{-4} to 10^{-1}. It can be seen that the channel noise raises the lower limit on the permissible channel noise compared with PCM. Therefore, with the channel capacity and transmitter power fixed, it is necessary to add error protection to the compressed transmitted data.

Figures 11-6, 11-7, and 11-8 illustrate the effectiveness of the (14, 10) and (7, 4) code at a channel bit error probability of 10^{-3} with the compression

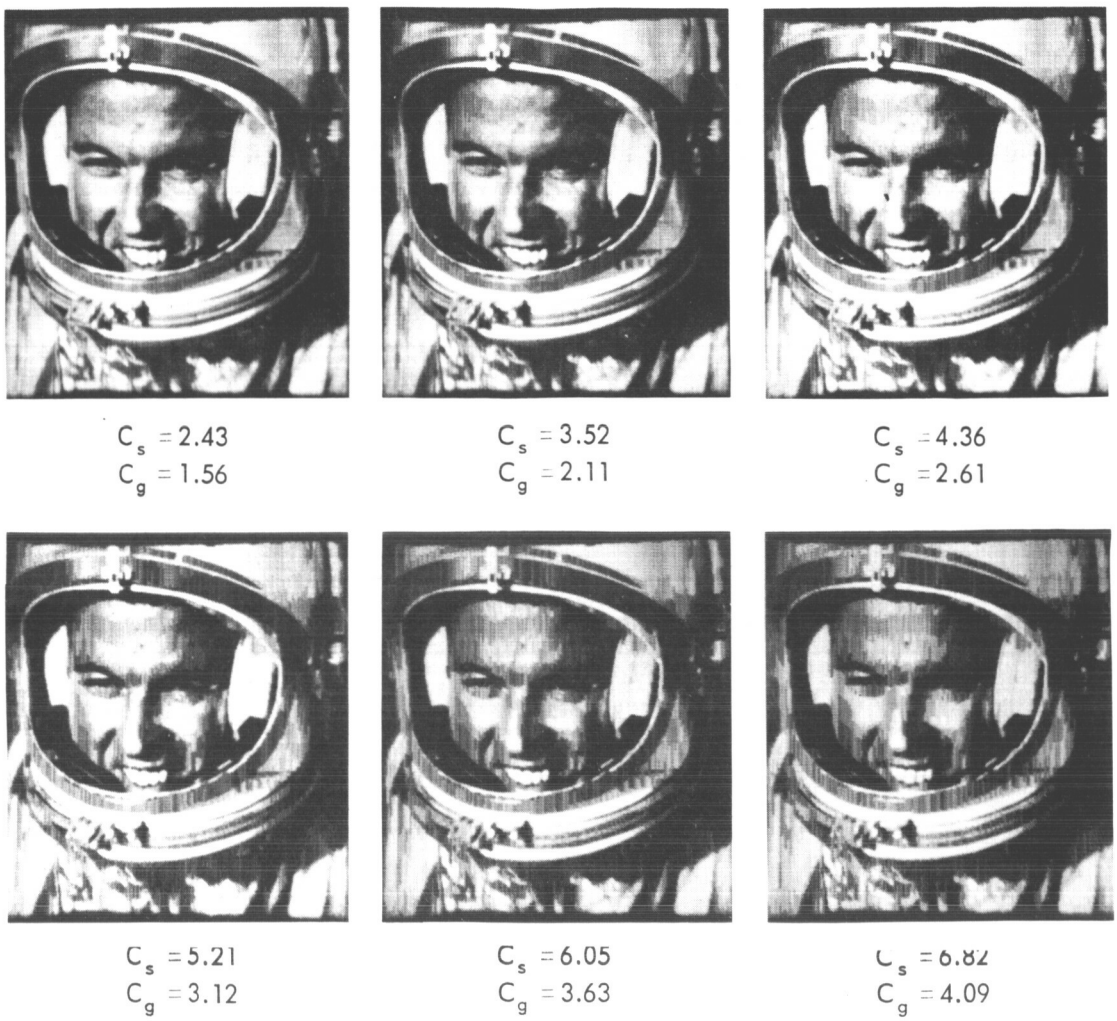


Figure 11-4. Subjective Comparison of the Effects of Zero-Order Prediction Data Compression

variable. Here the effects of channel noise become more pronounced as compression increases. However, the error protection is quite effective in reducing the error expansion.

RMS Error Results

As stated earlier, the rms error for the zero-order predictor can be considered as consisting of two parts: (1) The irreducible coding error and (2) the error caused by the effects of channel noise giving an rms error expansion. Figure 11-9 plots the rms error as a function of the gross average compression. At an error band of ±5 out of 64 possible levels, the rms error is 3.54 percent;

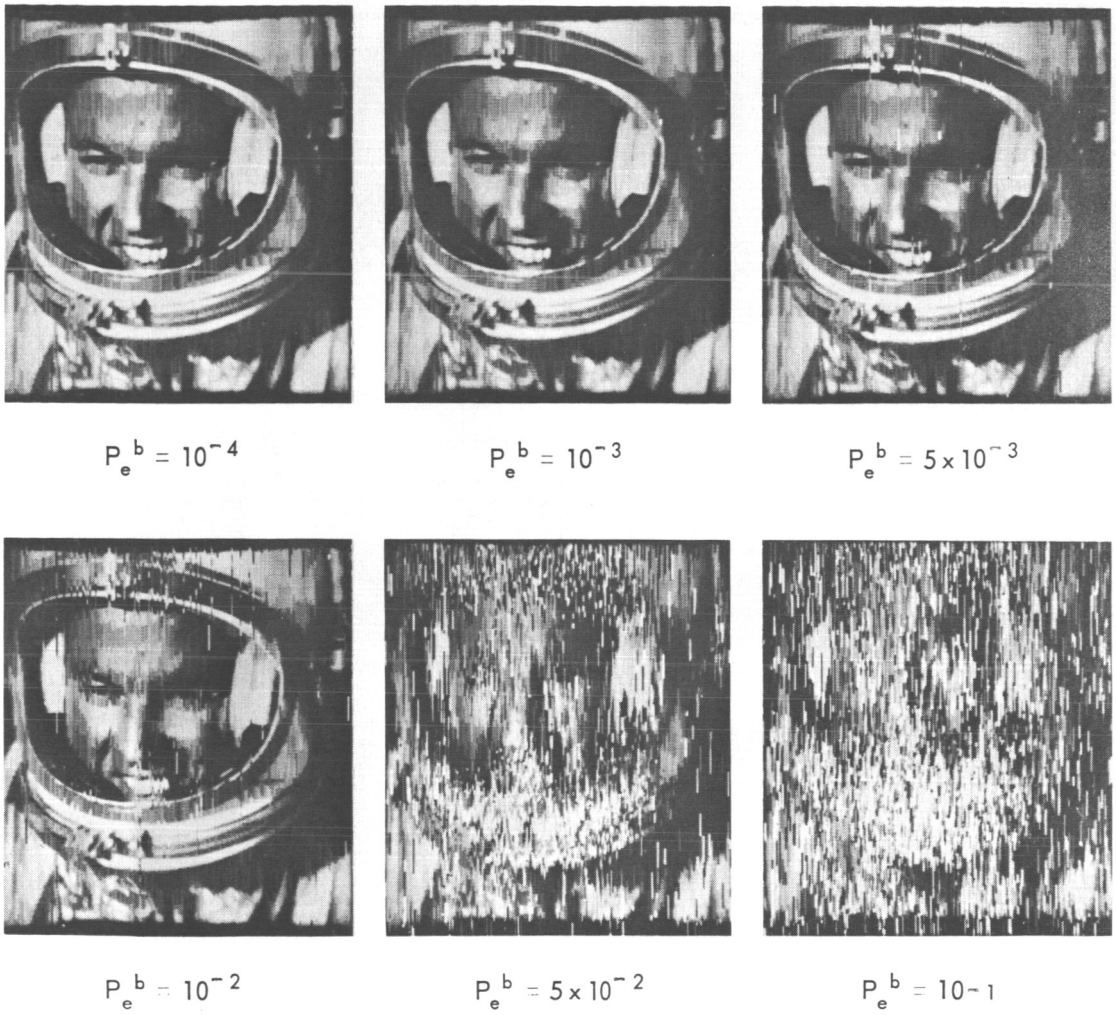
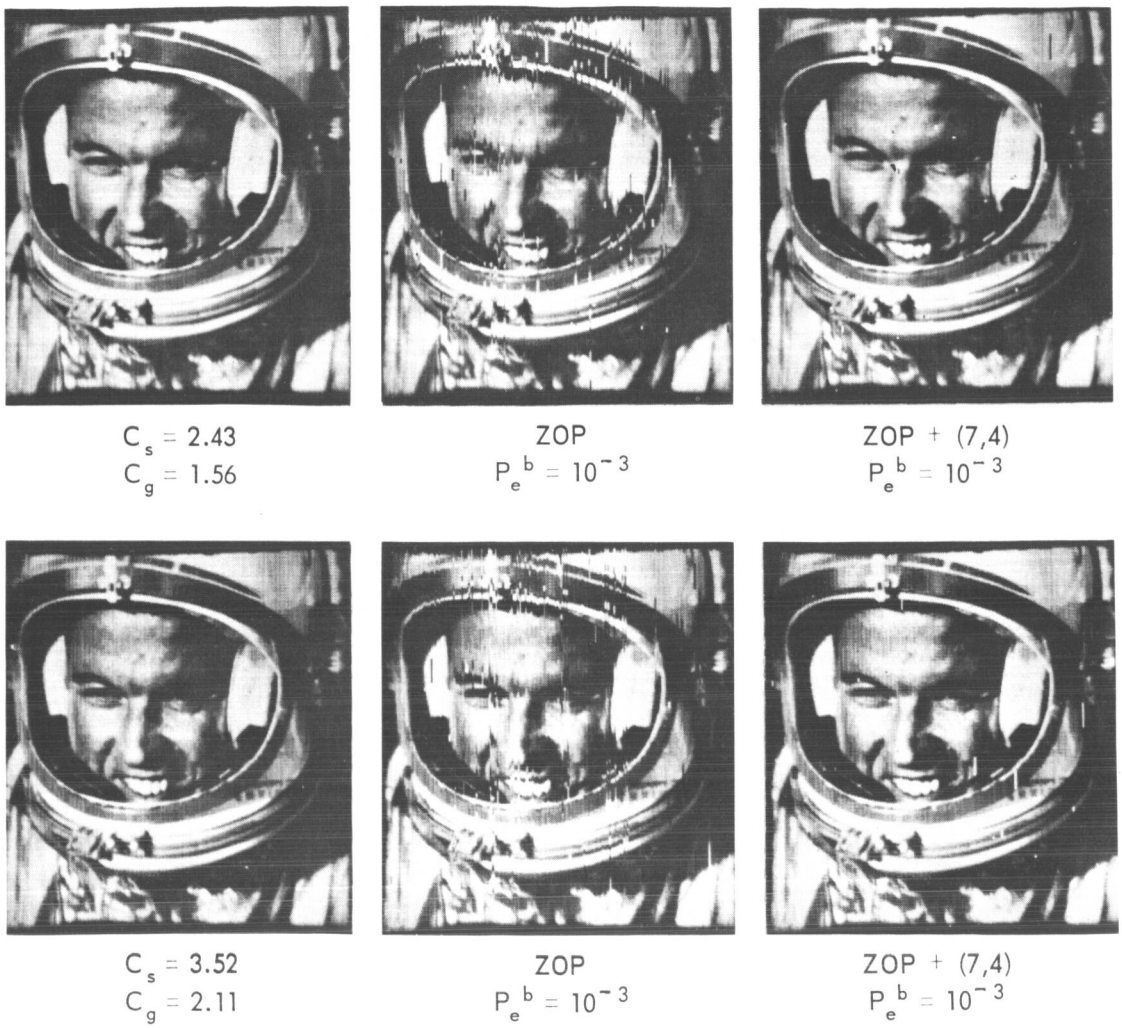


Figure 11-5. The Effects of Channel Noise on a Zero-Order Predictor with a Sample Reduction of 6:05 and a (14, 10) Hamming Code

however, the peak error is 7.85 percent. Since the peak error is normally specified and the actual rms error is less than one-half peak, it appears that peak error is not an especially good criterion in describing the zero-order predictor. Furthermore, recalling the subjective photos of Figure 11-3, which contain an rms error of 3.54 percent (a high value relative to the quantizing noise), it can be seen that subjectively the picture is acceptable. Therefore, it appears that neither the rms nor peak errors are acceptable error criteria for television data. Also given in Figure 11-9 is Equation 28 which is the analytical prediction of the rms error. It might be pointed out that although the rms error is not a good error criterion for TV in certain telemetry channels, it will be acceptable, and the analytical expression should be valid for telemetry data.



$C_s = 2.43$
$C_g = 1.56$

ZOP
$P_e^b = 10^{-3}$

ZOP + (7,4)
$P_e^b = 10^{-3}$

$C_s = 3.52$
$C_g = 2.11$

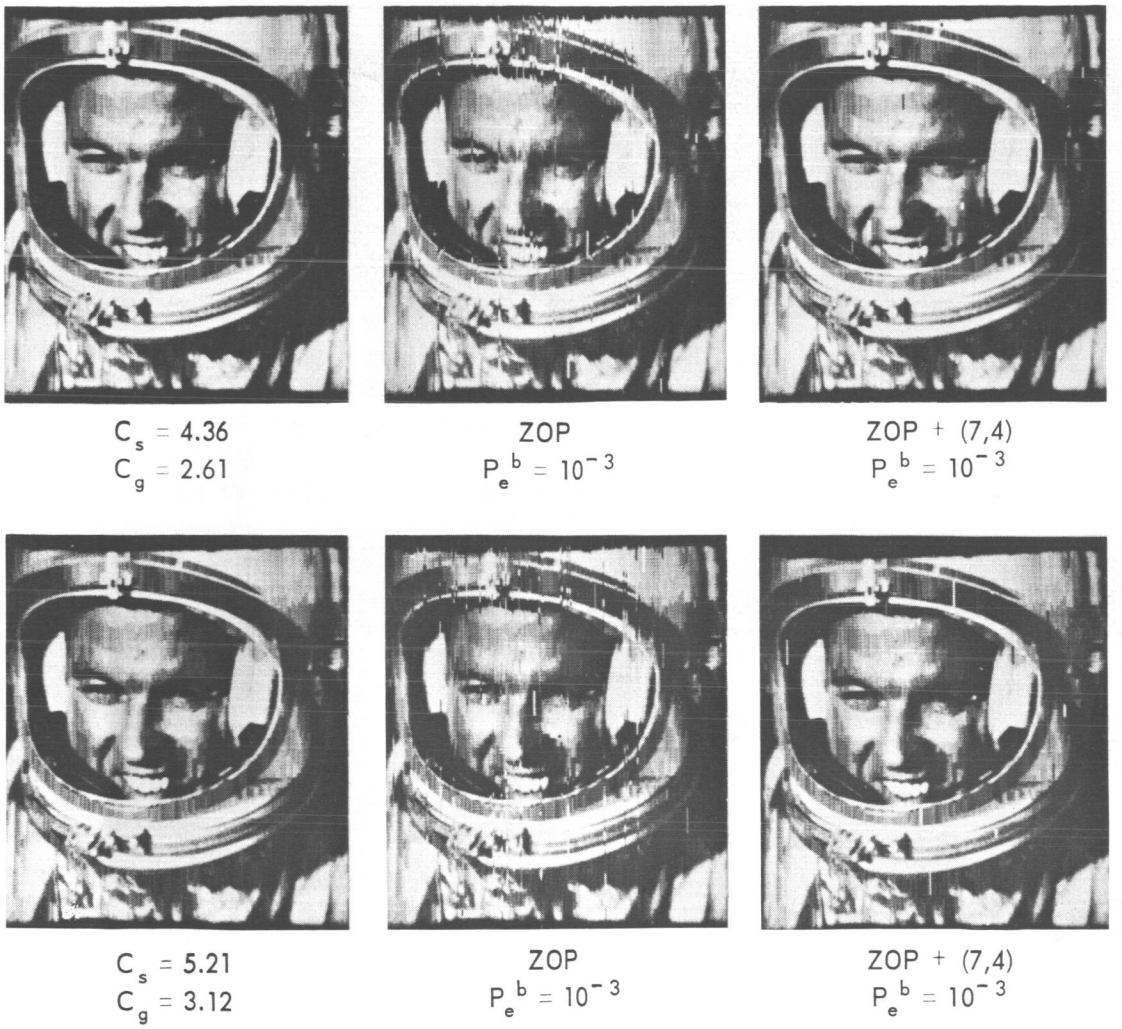
ZOP
$P_e^b = 10^{-3}$

ZOP + (7,4)
$P_e^b = 10^{-3}$

Figure 11-6. The Effects of Channel Noise on a Zero-Order Predictor Data Compactor

Figure 11-10 illustrates the effectiveness of the error-correction code along with the equivalent PCM error. The error-correction code, which is a monotonically increasing function with a decreasing bit error probability, can never reduce the error below the basic rms error of the coding technique. As the error rate increases, the rms error and the rms error with protection converge to approximately 25 percent. Below $P_e^b = 10^{-2}$, the effectiveness of the error-correcting code decreases, so its usefulness also decreases.

The analytical prediction based on Equation 29 agrees quite well with the experimental results. It appears that there exists an optimum error reduction as evidence by the PCM and zero-order predictor with the (14, 10) code having



$C_s = 4.36$
$C_g = 2.61$

ZOP
$P_e^b = 10^{-3}$

ZOP + (7,4)
$P_e^b = 10^{-3}$

$C_s = 5.21$
$C_g = 3.12$

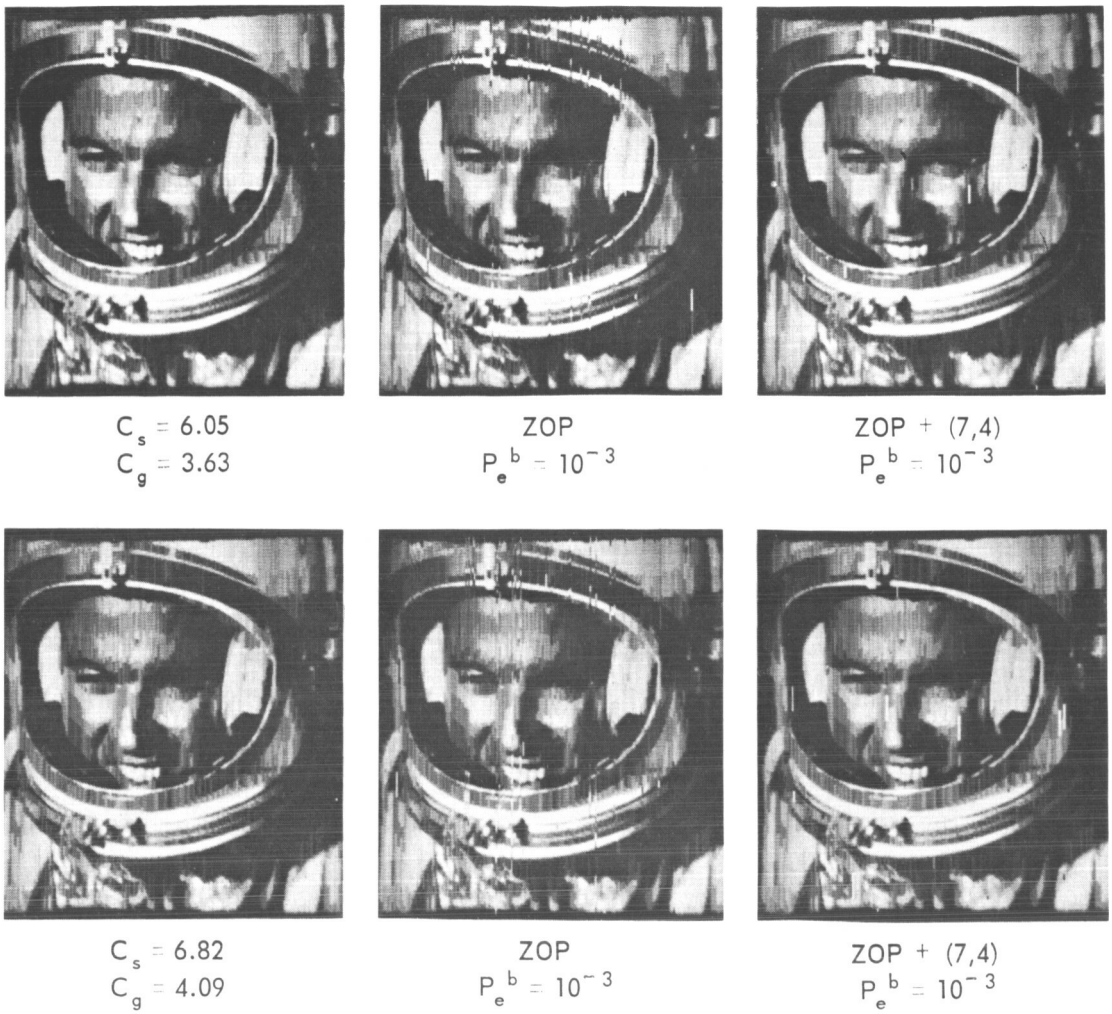
ZOP
$P_e^b = 10^{-3}$

ZOP + (7,4)
$P_e^b = 10^{-3}$

Figure 11-7. The Effects of Channel Noise on a Zero-Order Predictor Data Compactor

nearly the same rms error at $P_e^b = 10^{-2}$. In this case, at the same rms error, the zero-order predictor without error protection operates at $P_e^b = 9 \times 10^{-4}$. Hence, Figure 11-11 plots what can be called a bit error improvement factor, that is, the ratio of the bit error probability after coding to the bit error probability before coding the same rms error. For example, at $P_e^b = 10^{-2}$ with the (14, 10) code, for the same rms error without protection, the zero-order predictor operates at $P_e^b = 9 \times 10^{-4}$, or a bit error improvement of 11.1.

The improvement with coding can also be seen from Figure 11-12 which plots the difference between the rms error without and with error protection as a function of the channel bit error probability. Here, the error-correcting code's



$C_s = 6.05$
$C_g = 3.63$

ZOP
$P_e^b = 10^{-3}$

ZOP + (7,4)
$P_e^b = 10^{-3}$

$C_s = 6.82$
$C_g = 4.09$

ZOP
$P_e^b = 10^{-3}$

ZOP + (7,4)
$P_e^b = 10^{-3}$

Figure 11-8. The Effects of Channel Noise on a Zero-Order Predictor Data Compactor

figure of merit is multiplied by the rms error where both curves are a function of P_e^b. The figure of merit of the code is a monotonically increasing function of decreasing P_e^b whereas the rms error for the zero-order predictor is a monotonically decreasing function to a constant value with decreasing P_e^b. This seems to indicate that there exists an optimum operating channel bit error probability for maximum error reduction; however, this error probability is not necessarily optimum for minimum rms error. Hence, by plotting the difference of Equation 23 and the equation modified by the figure of merit of the error-correcting code, one can obtain Figure 11-12, which experimentally gives the optimum P_e^b for maximum rms error reduction.

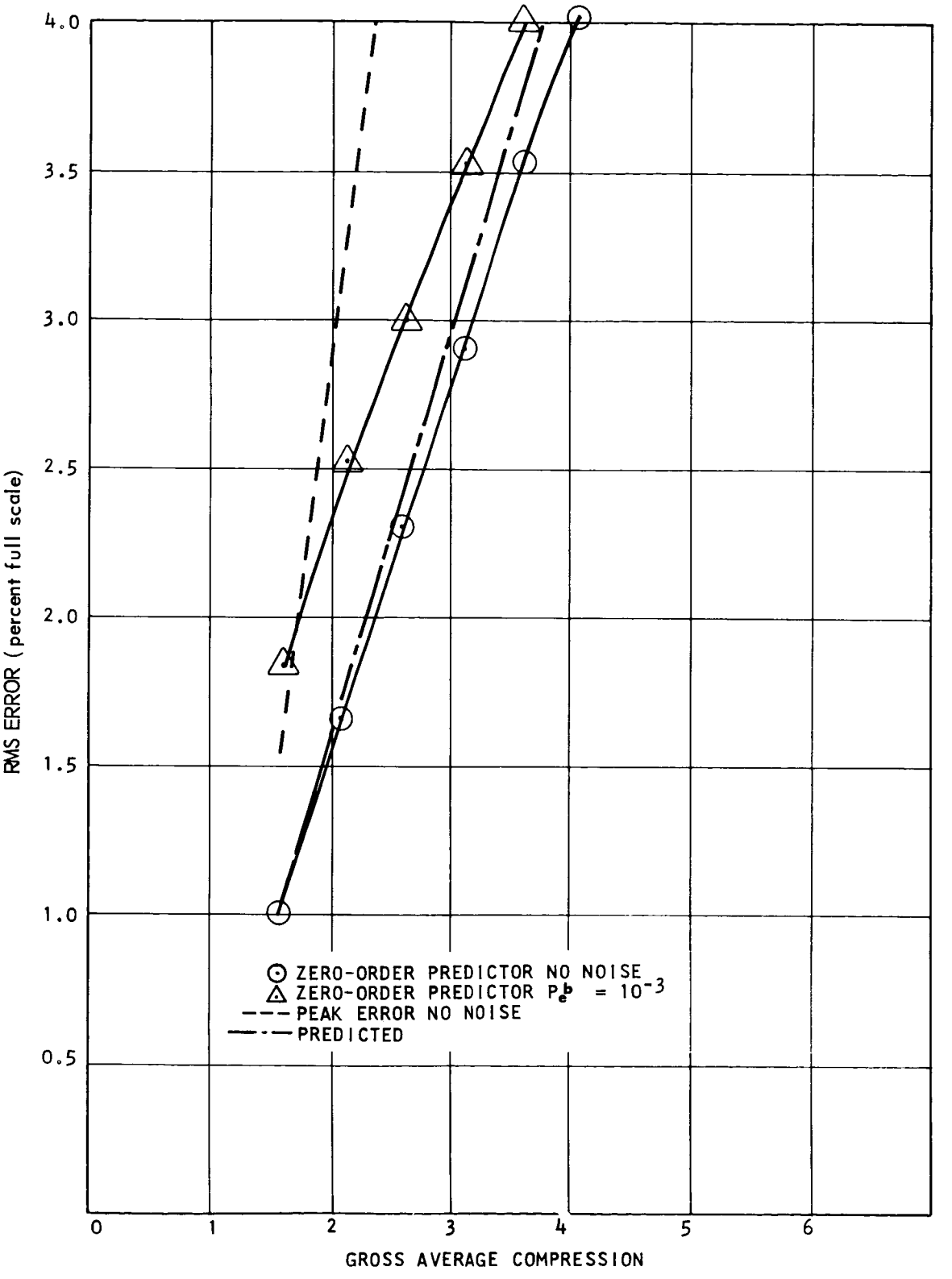


Figure 11-9. RMS Error Versus Gross Average Compression for Zero-Order Predictor

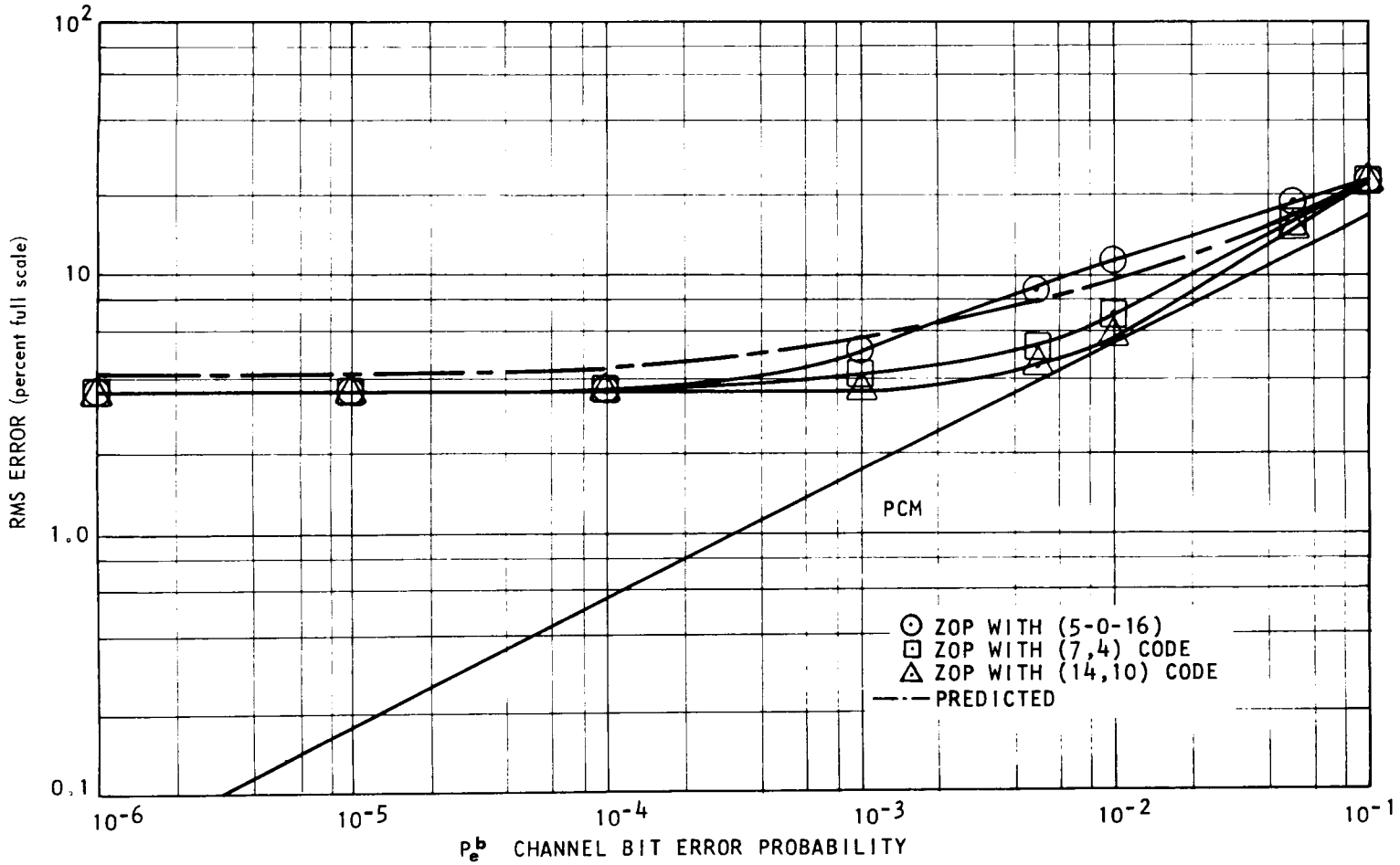


Figure 11-10. RMS Error Versus Channel Bit Error Probability for Zero-Order Predictor with and without Error Protection

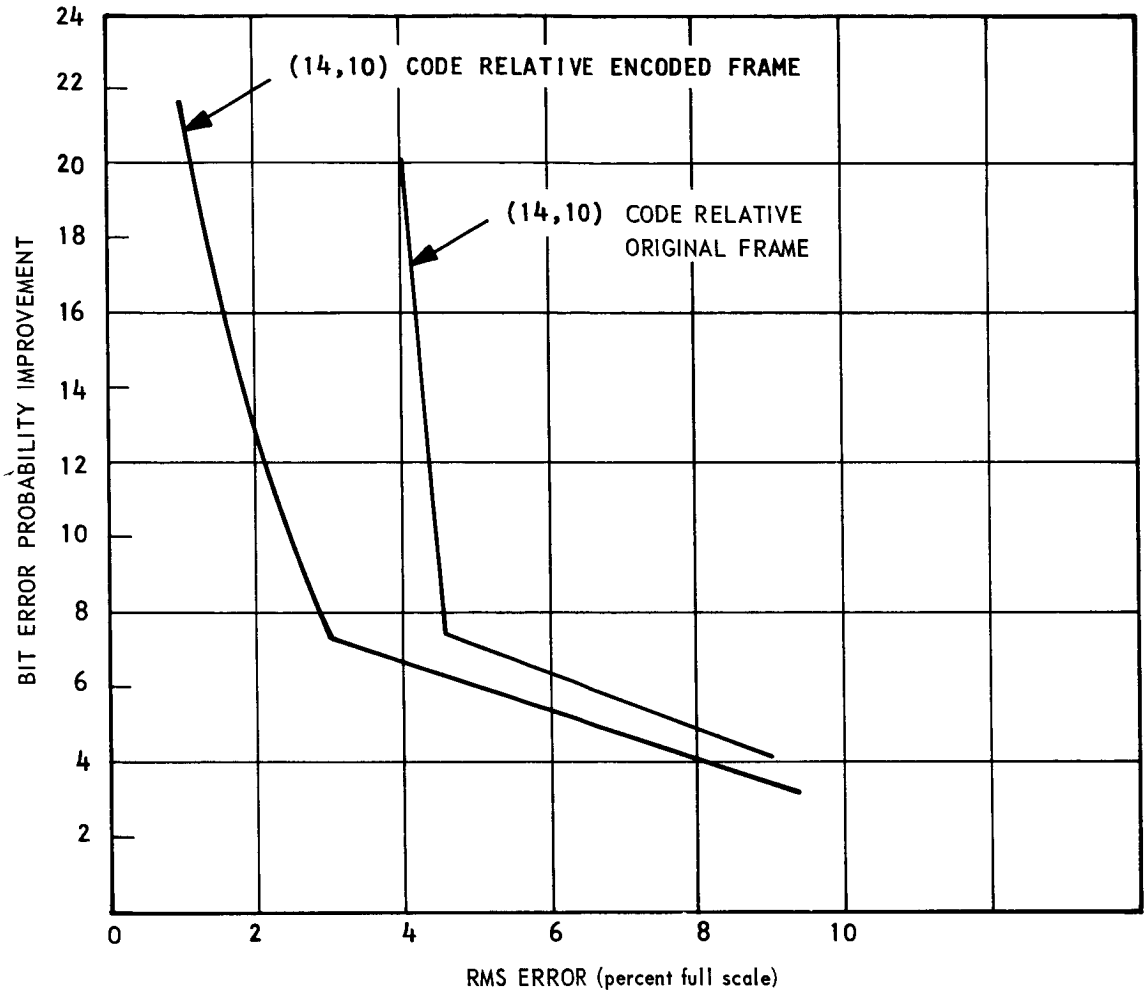


Figure 11-11. Bit Error Probability Improvement for (14, 10) Code Relative to Original and Encoded Frames for Zero-Order Predictor (5-0-16)

In view of Figure 11-12, the data were replotted in Figure 11-13 to give a figure of merit for the error-correcting code. Given a minimum acceptable rms error before error correction, it is possible to determine the rms error after error correction. This can be considered as a figure of merit for the error-correcting code with the zero-order predictor.

Figures 11-14 and 11-15 show how another error criterion could possibly be employed. Both rms error and percent similarity for the zero-order predictor and linear approximator relative to the encoded form are plotted. Note that for the zero-order predictor without channel noise the rms error is 3.54 percent whereas the percent similarity relative to the original data is 35.2 percent. It appears that the percent similarity has more meaning than the rms error at this particular point.

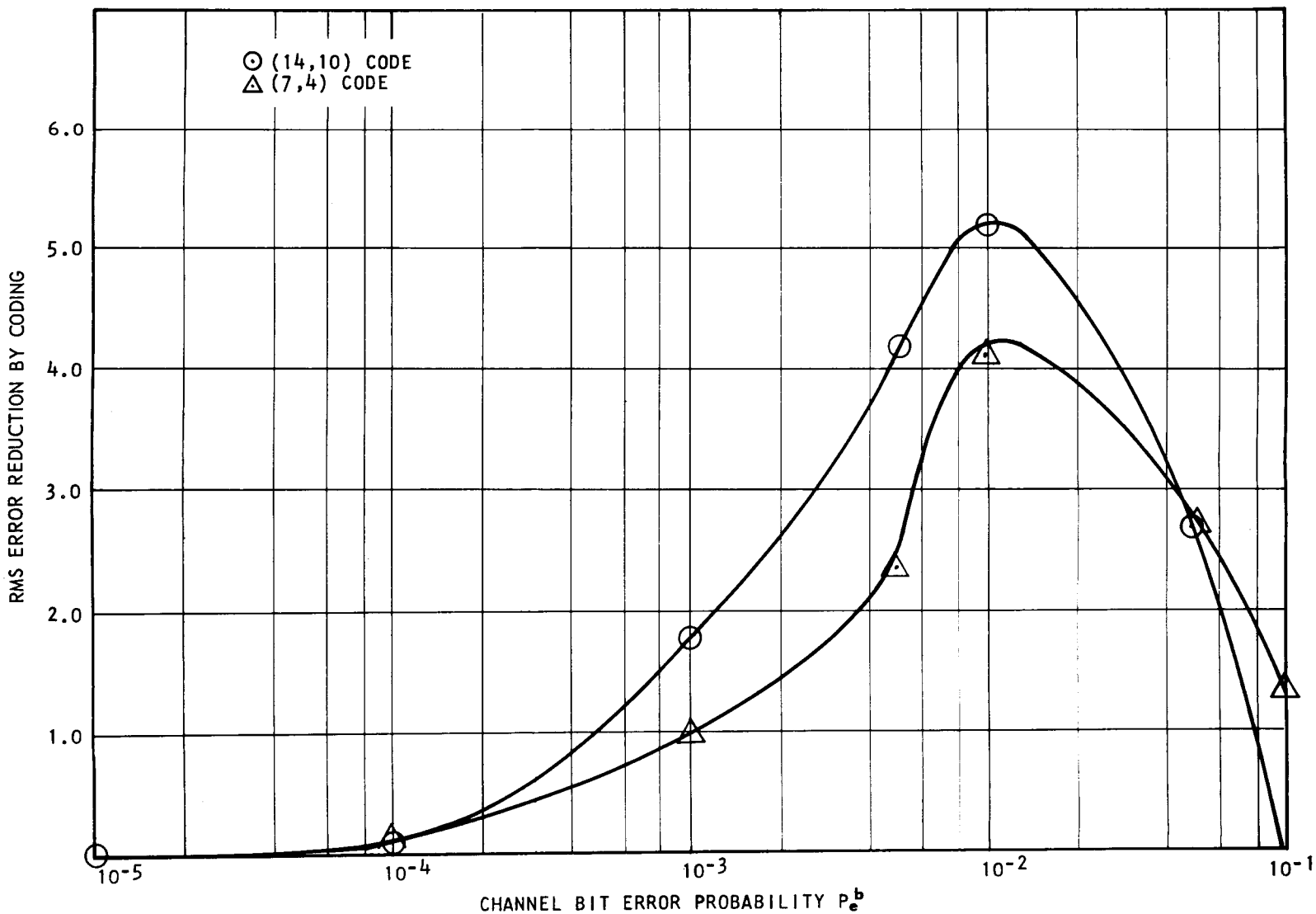


Figure 11-12. Reduction in RMS Error by Coding Versus Channel Bit Error Probability for the Zero-Order Predictor (5-0-16)

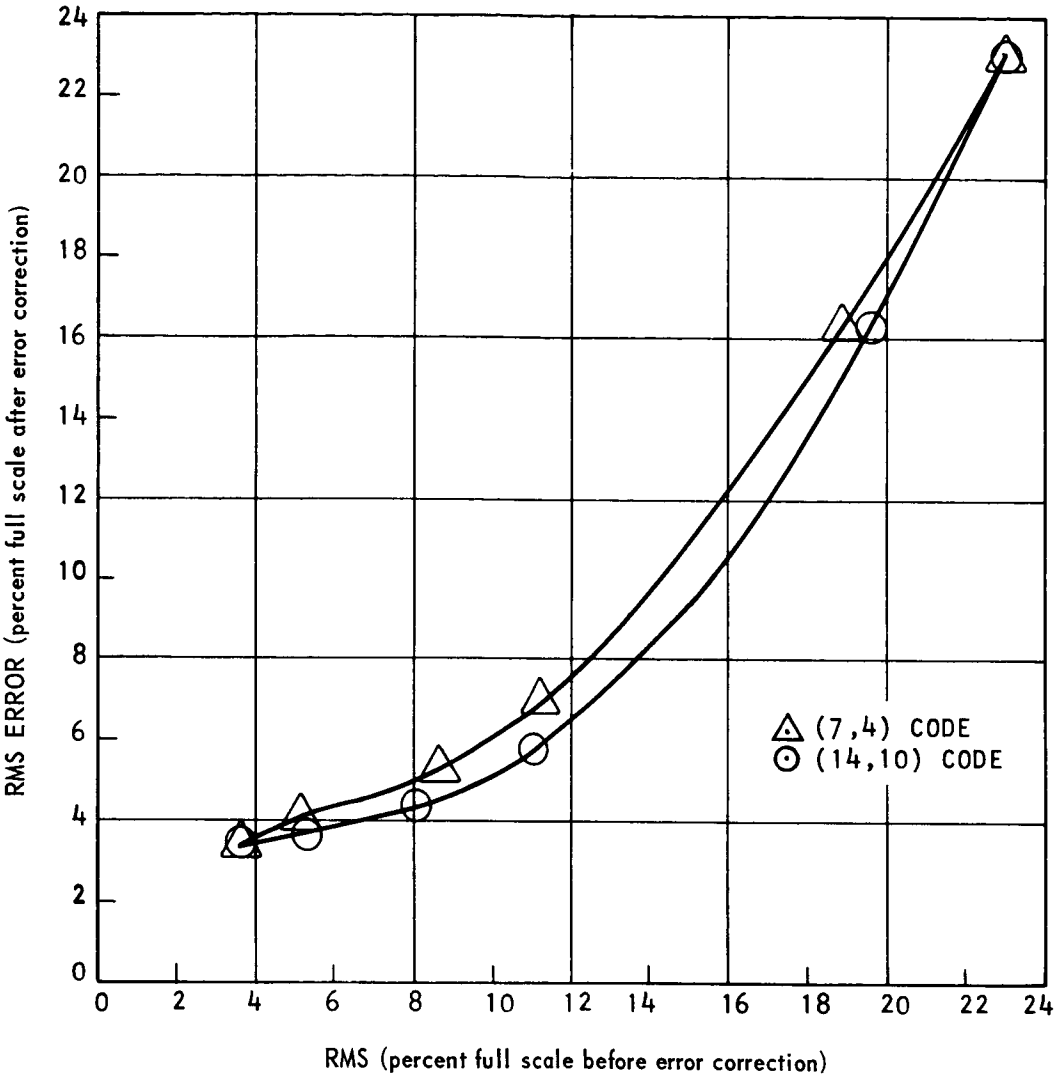


Figure 11-13. RMS Error After Error Correction Versus RMS Error Before Error Correction for Zero-Order Prediction (5-0-16)

CONCLUSIONS

It can be concluded that under the constraints of fixed transmitter power and fixed bandwidth, compression can be used to obtain more information per unit time. Furthermore, under these constraints, some form of error protection must be employed. The type and form of the error protection will be a compromise between implementation, desired efficiency, and channel characteristics. When error protection is employed on the class of techniques, the effectiveness of the code is decreased considerably. Sample reduction is not a valid way to compare system performance at this class of techniques. The effects of channel

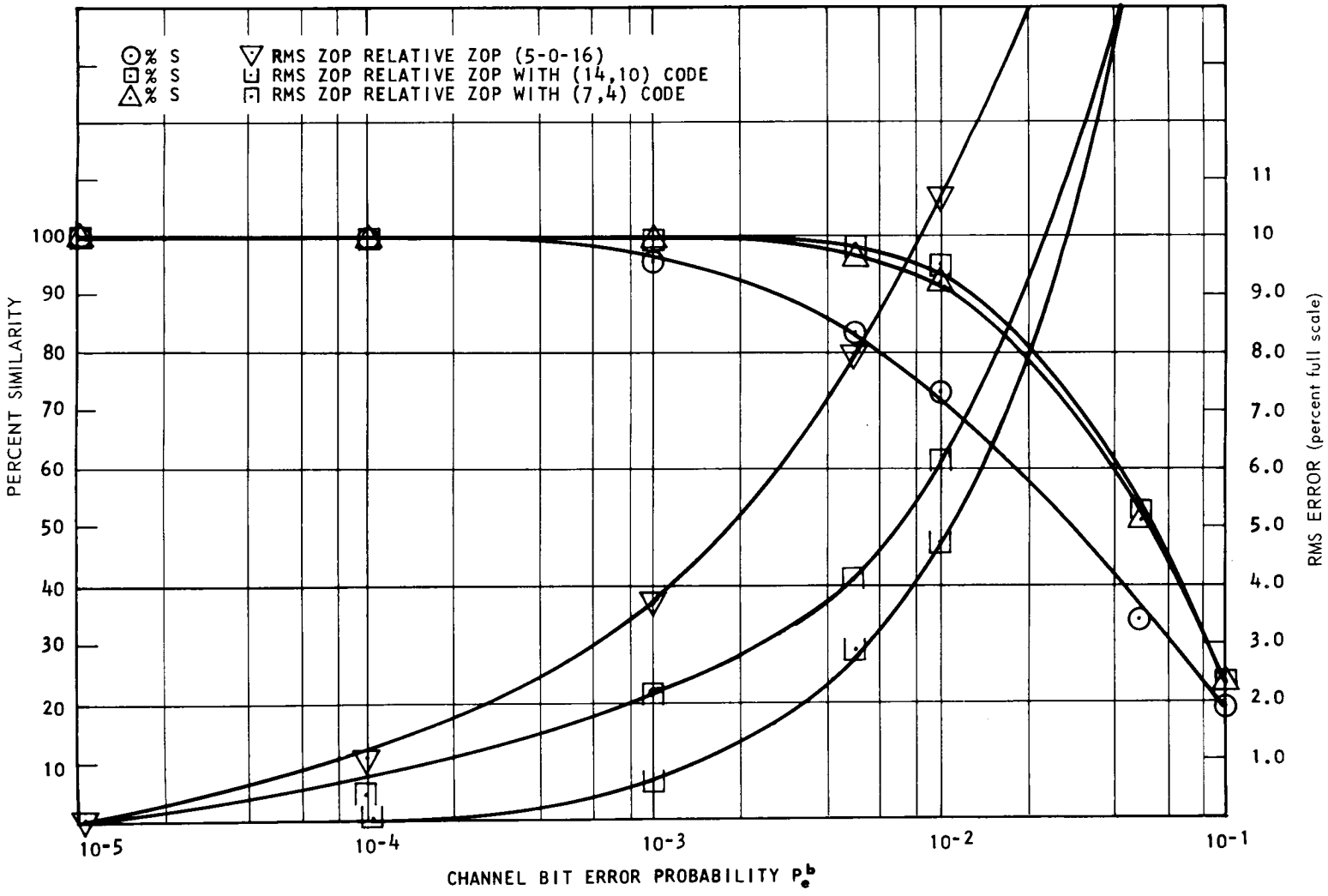


Figure 11-14. RMS Error and Percent Similarity Versus Channel Bit Error Probability for Zero-Order Predictor with and without Error Protection

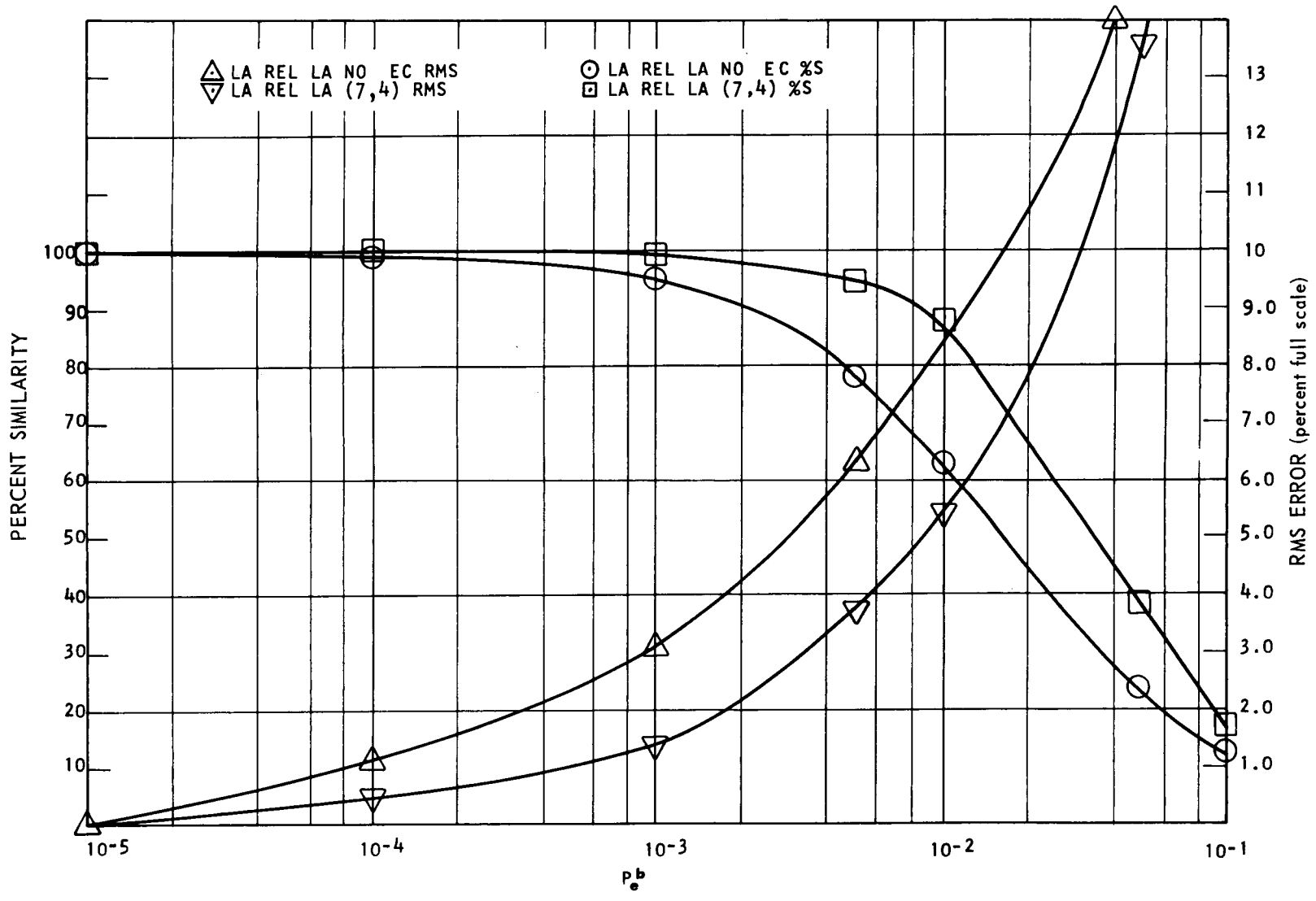


Figure 11-15. RMS Error and Percent Similarity Versus Channel Bit Error Probability for Linear Approximate with and without Error Protection

noise are more pronounced on compressed systems; therefore, error protection is required. When addressing of the nonredundant samples and the redundancy of an error-protection code are considered, the net coding compression is reduced more than 50 percent.

This particular experimental program requires additional effort to verify the analytical expressions as applied to the broad category of telemetry data. Various addressing techniques, including optimum encoding of the addresses, might be investigated as a means of increasing the net compression and the effects of channel noise on the self-synchronizing codes to be investigated. Although the television source was employed in the analysis, it must be emphasized that, in general, these results will apply to the broad class of adaptive sampling telemetry techniques since the television source can be considered to be an extremely active telemetry channel.

REFERENCES

1. Zukerman, L. G., and Ross, L., "Some Coding Concepts to Conserve Bandwidth," Proc. National Telemetering Conf. (May 24-27, 1959, Denver, Colo.), pp. 138-142.

2. Hulme, J. R., and Shomburg, R. A., "A Data Bandwidth Compression for Space Vehicle Telemetry," Proc. National Telemetering Conf. (May 23-25, 1962, Washington, D.C.), New York, N. Y.: The Institute of Radio Engineers, Vol. 1, Paper 3-2, p. 14.

3. Webber, D. R., and Wynhoff, F. J., "The Concept of Self-Adaptive Data Compression," Proc. National Symposium on Space Electronics and Telemetry (October 2-4, 1962, Miami Beach, Fla.), Paper 4.1, p. 10.

4. Berkowitz, M., "Adaptive Data Sampling System," National Space Electronics Symposium (October 1-4, 1963, Miami Beach, Fla.), Paper 4.1, p. 20.

5. Morton, W., and McLeannan, M. A., "A Prototype Selective Monitoring System for Physiological Data." National Space Electronics Symposium, (October 1-4, 1963, Miami Beach, Fla.), Paper 8.2, p. 20.

6. Blasbalg, H., and Van Blerkom, R., "Message Compression," The Institute of Radio Engineers Transactions on Space Electronics and Telemetry, Set-8 (3):228-238, September 1962.

7. Eisenberger and Posner, "Data Compression by Quantized Floating Aperture System," JPL Space Programs Summary No. 37-17, October 1962.

8. Gardenhire, L. W., "The Use of Digital Data Systems," National Telemetering Conference Proceedings (May 20-22, 1963, Albuquerque, N. M.), Paper 3-1, p. 8.

9. Palermo, C. J., Palermo, R. V., and Horwitz, H., "The Use of Data Omission for Redundancy Removal," Int. Space Electronics Symposium (November 2-4, 1965, Miami Beach, Fla.), pp. 11D1-11D16.

10. Raga, G. L., "A Unified Approach to Digital Television Compression." Proc. National Telemetering Conf. (April 13-15, 1965, Houston, Tex.), Paper TA 1.4, pp. 29-34.

12. ELIMINATING REDUNDANCY IN FIXED FRAME TELEVISION

L. W. Gardenhire

Radiation, Inc.
Melbourne, Florida

INTRODUCTION

The present author has been appalled by the disregard for the basic information content of original data being processed by various data compressors and also dismayed by the lack of practical considerations for what the limiting factors really are in areas such as conversion and channel noise. Rarely has the problem been related to how many uniformly spaced samples are needed to reproduce the original data; this is a necessary first step. The magic term "compression ratio" can become anything we want, depending upon the choice of the sampling rate. One can make any redundance reduction scheme appear better than any other scheme by merely undersampling or oversampling.

Any digital redundancy reduction scheme that has ever been devised is directly related to how the waveform was originally sampled. The tremendous amount of redundancy occurs because the data are not Gaussian or stationary. The chosen sampling rate is based upon the frequency of the fastest transient expected; most of the time these frequencies do not occur, and therefore there is much redundancy in the data. The sampling rate must be high, however, in order to capture the waveforms in case they do occur. Thus, there is a chance to provide a real adaptive system that will not lose data but will reconstruct any waveform with a minimum number of samples to any desired accuracy specified by the user.

SPECTRAL CHARACTERISTICS

There is a direct relationship existing between the spectral characteristics of the original waveform that has been sampled and the interpolation error created by whatever interpolation method is used to reconstruct the sampled data. This holds for the output of a transducer, response of any order system to a step function, reconstruction of one line of TV data, or, for that matter, for the average spectral characteristics of an entire TV picture.

The practical optimum interpolation process for any data that have been properly sampled is one order less than the order of the original data. When the

data are undersampled, this relationship does not hold and has led many people to assume that zero-order redundancy schemes are more effective than higher order ones. By the same token, many reports have been written where large unrealistic compression ratios have been shown simply by oversampling the data. Most of these reports do not show any basis for selecting the sampling rate or even mention what it was. When the original data are not sampled fast enough to recover them to a given accuracy, it is not realistic to try to reconstruct them to a greater accuracy. It is true that slight gain is realized by going to higher order interpolation schemes, but the gain is very slight and the interpolation error never goes to zero.

REDUNDANCY REDUCTION

Redundancy reduction hardware does not need to be complicated. The only function of predictors and interpolators is to reconstruct the waveform to a given accuracy with a minimum number of samples. In other words, the waveform is sampled fast enough at a uniform rate to assure that the sample falls at the proper place on the waveform; then interpolation can be performed between these nonredundant samples and checked to see that they are within the desired accuracy. Any redundancy reduction scheme reconstructs the data and, in turn, produces all of the errors created by the original sampling plus its own reconstruction errors. If the data were undersampled in the first place and caused aliasing errors or foldover frequencies that do not exist, the redundancy scheme cannot tell them from real data and thus transmits them as nonredundant data.

Because a relationship exists between the spectral characteristics of the original waveform and the interpolation process, there is also a relationship to the amount of redundancy or to the redundancy curve for any reduction process. Since any redundancy reduction scheme is nothing more than a reconstruction or interpolation process with a minimum number of samples, it holds that the that the best redundancy reduction scheme is also one order less than the order of the data.

STUDY PROGRAM

Several redundancy reduction studies have been performed on fixed frame TV pictures after they have been slowly scanned, and then transmitted at a faster rate over analog circuits. These pictures, such as Tiros and Nimbus APT, have a large amount of conversion noise, which is the limiting factor in how much reduction is possible. The scanning process itself is rather noisy, particularly for slow electronic scanning systems. A 20 db signal-to-noise ratio at the output of a Vidicon is considered good. To transmit a six-bit picture with no reduction, one needs about 40 db of signal-to-noise-ratio. In order to avoid the conversion

noise problem as much as possible, a flying spot scanner was used in this study to scan high quality photographs.

The redundancy studies reported here were performed with special Radiation Inc., equipment called a "Data Management Analyzer." The name is based upon an understanding of the relationships mentioned above which allows it to be used to determine much about the characteristics of data (i.e., what the proper sampling rate is if the data are analog, or what the interpolation error is if they are digital). Its operational speed is equal to two computers with an access time of 1.3 microseconds. It is a redundancy remover capable of either a zero-order prediction or the "fan method" patented by Radiation Inc., which is a first-order predictor and interpolator. It also contains a reconstructor that is capable of either zero- or first-order reconstruction.

DIGITAL PICTURE TRANSMISSION

A great deal has been written about digital picture transmission; however, its actual use has been very limited because of the added complexities of both the transmitting and receiving station as well as the increased bandwidth requirements. Redundancy reduction techniques have advanced to the point where the bandwidth requirements can be greatly reduced. Integrated circuits can cut both the cost and complexity.

When one considers the extreme bandwidth requirements needed for digital picture transmission, it is quite obvious that here is one of the best places to apply redundancy reduction techniques first. Space probes with limited power available and some security requirements make digital TV a must in many cases. As discussed earlier, the first and foremost item in any digital picture transmission system is to assure that the sampling rate has been chosen to provide a resolution equivalent to the basic system capability. If the rate is higher than this, system capability is wasted; and, if it is lower, the resolution of the system is decreased.

Although one picture may appear to the eye to be adequately sampled, the next picture with the same sampling rate, having more fine detail, will be quite obviously undersampled. A picture that is undersampled may also look quite good before the redundancy removal process; however, after it has been reduced, the undersampling damages the picture much more than an adequately sampled picture reduced to the same number of nonredundant samples. This shows the importance of having a sample fall at the correct place on the waveform produced by scanning.

CHOICE OF SAMPLING RATE

As was the case in PCM telemetry, rules of thumb for selecting sampling rates have been generated for digital TV. Just as five samples per cycle of the maximum frequency has no basis, having as many samples in a line as there are lines in a square picture has no basis. This would hold only in an ideal case where the picture was scanned by a square spot that stepped across the picture with a spot size equal to the resolution elements. When a square spot is used to scan across a picture with maximum resolution elements at a fixed rate, triangular waveforms are produced that are very high in harmonic content. When the spot is round and smaller than the resolution element, square waves are produced, which are even higher in harmonic content. These frequencies, when added to and subtracted from the sampling frequency, produce foldover frequencies that are very detrimental if the sampling rate is not high enough.

To show this effect, a standard test pattern was scanned with a 300-line/inch scanning system. This analog pattern, as shown in Figure 12-1, has been transmitted and reconstructed. As can be seen by examining the wedge of converging lines, the resolution is not 0.003 inch (300 lines/inch) but more nearly 0.005 inch (200 lines/inch). This can be seen by noting where the lines merge.

By assuming that the analog voltage produced by the scanning process is sampled at 250 samples per line, encoded to five bits, and then reconstructed, the results are as shown in Figure 12-2. Note the extreme interference going out to less than 0.025 inch (50 lines/inch). As the sampling rate is increased, the interference patterns move towards the center as seen in Figure 12-3. This is an enlargement of only the center circle or 150 lines to the inch. The sampling rate has been increased to 750 samples per line. Note the interference patterns just inside the 150-line/inch circle.

When the rate is increased to 1000 samples per line, the fringes move in to around 200 lines to the inch as seen in Figure 12-4. One thousand samples per line is then the proper sampling rate to maintain the resolution of this particular system.

The rate was also checked by studying the spectral characteristics of several pictures (sampled 1000 samples per line); the rms interpolation error varied between 3 and 7 percent.

TEST PORTRAIT

Although most of the studies performed have been on cloud cover pictures taken by the astronauts, the pictures shown here are more pleasing to look at,

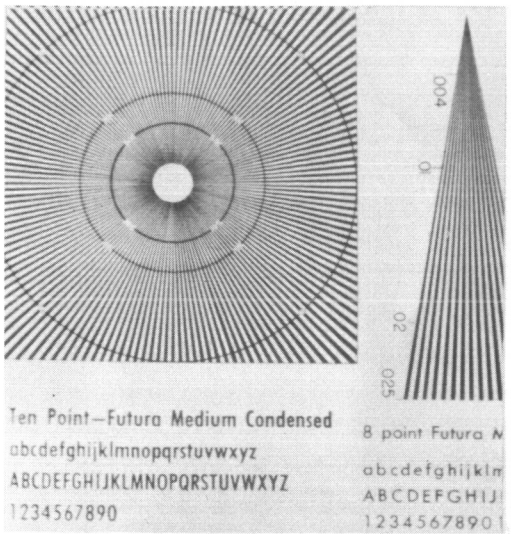


Figure 12-1. Analog Test Pattern, 300 Lines/Inch

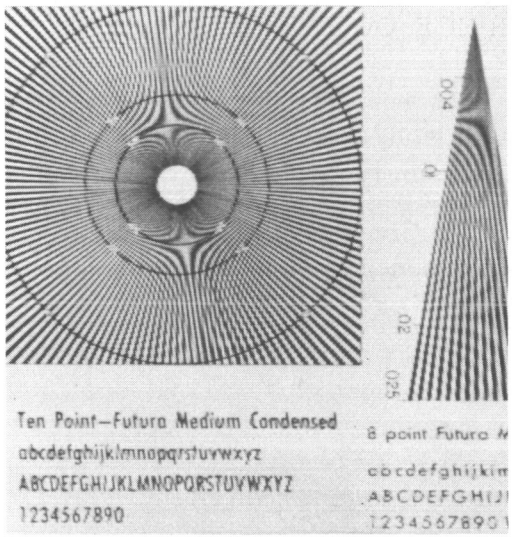


Figure 12-2. Digital Test Pattern, 200 Samples/Line 216,000 Total Samples

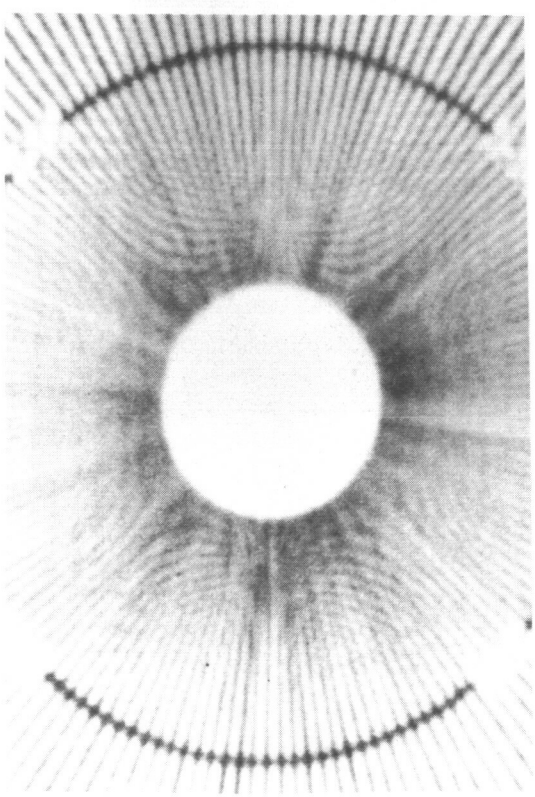


Figure 12-3. Digital Test Pattern, 750 Samples/Line 810,000 Total Samples

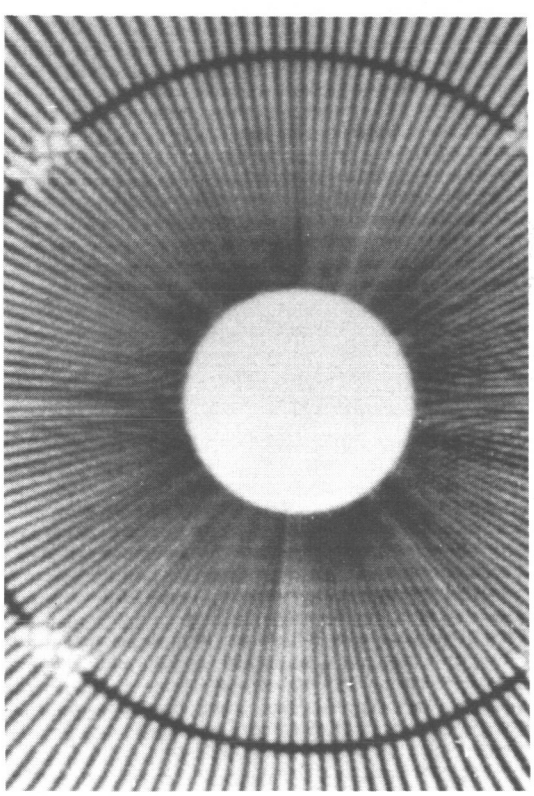
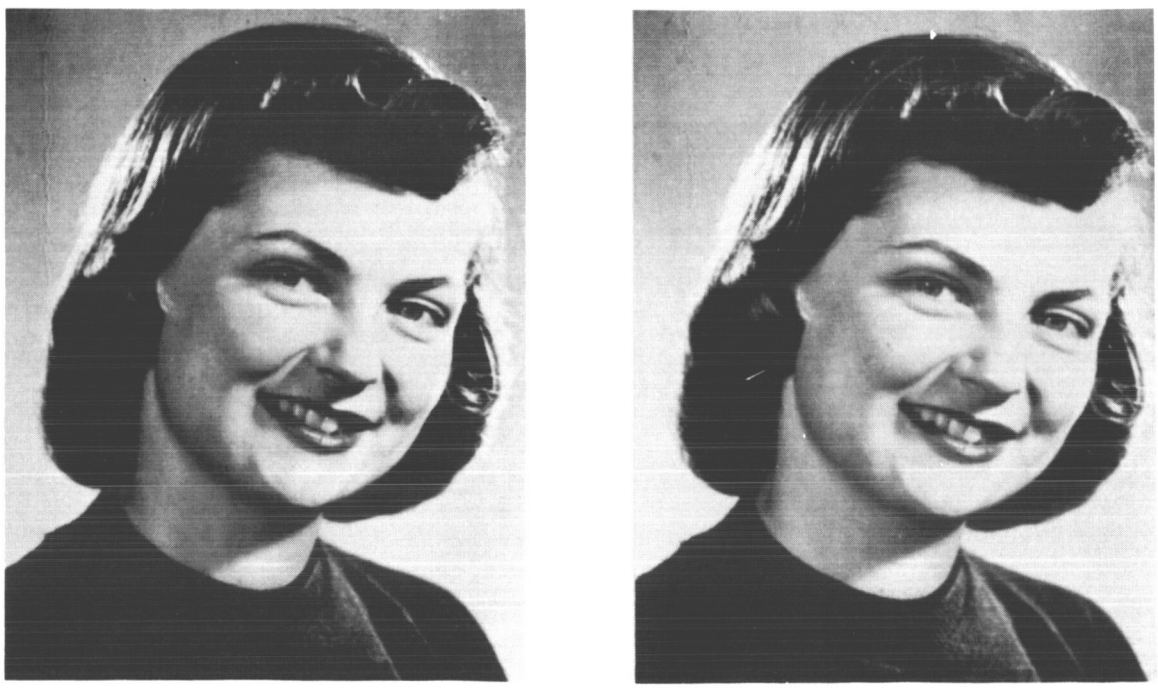


Figure 12-4. Digital Test Pattern, 1000 Samples/Line 1,080,000 Total Samples

and the effects of the redundancy removal process is much easier to observe on pictures that we are used to.

Figure 12-5 shows a portrait that has been scanned and reproduced using analog means on the left and digital means on the right. The original picture was scanned with a 0.003 inch spot using a photomultiplier tube. The resolution was 300 lines to the inch. The picture size was 2.875 × 3.60 inches. The picture is composed of 1080 lines; therefore, the digital picture has 1,080,000 five-bit samples or 5,400,000 bits.



ANALOG PICTURE
300 LINES/INCH
1080 LINES - 2.875 INCHES

DIGITAL PICTURE
1000 - 5 BIT SAMPLES/LINE
1,080,000 SAMPLES
5,400,000 BITS

Figure 12-5. Comparison of Analog and Digital Picture

In the original pictures, it is difficult, if not impossible, to tell the two apart. Much has been lost in the photographic reproductions and the differences are attributed more to contrast variations in the photographic processing than in the digital process.

RESULTS OF FAN METHOD WITH A FIRST-ORDER PREDICTOR AND INTERPOLATOR

The results of processing the digital picture (Figure 12-6(a)) by the fan redundancy reduction method using a tolerance of 3.1 percent, can be seen in Figure 12-6(b). The analog voltage from the photomultiplier tube was sampled to eight bits, the redundancy removed, and then only the five most significant bits transmitted. This was done in order to keep the quantization noise small enough not to affect the redundancy reduction process. The 3.1 percent tolerance then means that the picture is reproduced to within 8 levels in a full scale of 256 levels. In actuality, the reproducing device is only capable of something less than 16 levels of gray. This then means the tolerance or peak error is within one-half gray level of the original, which is less than the resolution of the reproducer. The 1,080,000 original samples in the picture have been reduced to 172,000 total samples or 159 samples per line. This yields a sample reduction S_r of 6.28:1.

Some method must be used to determine where the nonredundant samples occur; this will require the transmission of additional information. In this case, a four-bit run length code was used. The four-bit code following each five-bit nonredundant data sample tells how many redundant samples have been dropped since the last transmitted sample. If the lengths exceed 16, the sixteenth sample is transmitted even though it may be redundant. This means that the bit reduction B_r is five-ninths of the sample reduction S_r, or 3.49:1 in this case.

Again it is virtually impossible to tell the pictures apart. If it were not for conversion noise present on the data to be processed, the tolerance could be extended to 1 part in 16 or 6.25 percent, and the two pictures would be almost identical.

RESULTS OF FAN METHOD USING 6.25 PERCENT TOLERANCE

When the tolerance is extended to 6.25 percent, the results can be seen in Figure 12-6(c). Here, the original 1,080,000 samples have been reduced to 98,000; only an average of 91 samples per line is being transmitted. This gives a sample reduction S_r of 11.02:1 and bit reduction B_r of 6.12:1. The smearing effect is due to not reproducing the original waveform; this is largely due to the conversion noise present. This is not surprising when one realizes that only 9 percent of the original samples are being used to reconstruct the picture.

RESULTS WITH FIVE-BIT TIMING WORD

When the reduction gets large, the number of times the run length code is exceeded becomes greater and greater. There is a break point where a five-bit

A

DIGITAL PICTURE
1,080,000 TOTAL SAMPLES
1000 - 5 BIT SAMPLES/LINE

B

3.1 PERCENT TOLERANCE
172,000 - 9 BIT SAMPLES
159 SAMPLES/LINE
15.9 PERCENT OF TOTAL SAMPLES
S_r = 6.28:1
B_r = 3.49:1

C

6.25 PERCENT TOLERANCE
98,000 - 9 BIT SAMPLES
91 SAMPLES/LINE
9.1 PERCENT OF TOTAL SAMPLES
S_r = 11.02:1
B_r = 6.12:1

D

9.4 PERCENT TOLERANCE
57,000 - 10 BIT SAMPLES
53 SAMPLES/LINE
5.3 PERCENT OF TOTAL SAMPLES
S_r = 18.95:1
B_r = 9.47:1

Figure 12-6. Reduction By Fan Method

run length code will take fewer bits than will transmitting more four-bit samples. Figure 12-6(d) shows this effect. The tolerance has been extended to 9.4 percent or 24 levels in 256. The 1,080,000 samples have now been reduced to 57,000 nonredundant samples. This means that only 5.3 percent of the original samples are used to reconstruct the picture. These 57,000 samples are 10 bits long with 5 information bits and 5 timing bits. If a four-bit timing word were used, there would be 90,000 nine-bit nonredundant samples or 810,000 total bits. The picture with five-bit timing words requires only 570,000 total bits. Figure 12-6(d) has an average of only 53 samples per line. The sample reduction S_r is 18.95:1 and bit reduction B_r is 9.47:1. The smearing effect is much more pronounced in this case; however, the picture is still quite useable. This smearing is again the result of the conversion noise present in the digital data. It comes from the reconstruction of noisy samples that do not reproduce the same waveform as when a great many samples are used. Greater reduction can be obtained with less smearing, if this noise is reduced, either before or after sampling.

SPECTRAL CHARACTERISTICS OF TEST PORTRAIT

As mentioned earlier, the most effective redundancy reduction method is one which is one order less than the order of the data. The spectral characteristics of this picture were made by using a narrow-band spectrum analyzer and averaging the output over the entire picture. The results showed the picture to be second order. That is, the spectral response curve of the data broke at about 100 cycles and decayed at 12 db per octave.

RESULTS OF STEP METHOD USING A ZERO-ORDER PREDICTOR

Figure 12-7(a) to 12-7(f) shows the operation of a zero-order (step method) redundancy reduction method on the same test picture. Figure 12-7(b) is for a tolerance of 3.1 percent or 8 levels out of 256. The 1,080,000 original samples have been reduced to 230,000 nonredundant samples, or 213 samples per line. The sample reduction S_r is 4.70:1, and bit reduction B_r is 2.61:1. The fan method required 172,000 samples at this tolerance.

Figure 12-7(c) is for a tolerance of 6.25 percent or 16 levels out of 256. The 1,080,000 original samples have been reduced to 123,000, or 114 nine-bit samples per line. The sample reduction S_r is 8.78:1, and bit reduction B_r is 4.88:1. The quality of the picture is quite good although false contouring is beginning to show. When the tolerance is increased so that the nonredundant samples are reduced to 100,000, as was the case for the fan method at 6.25 percent, the false contouring is quite bad. This can be seen in Figure 12-7(d), where the picture has been reduced to 102,000 nonredundant samples using a tolerance of 9.4 percent.

A

DIGITAL PICTURE
1,080,000 TOTAL SAMPLES
1000 – 5 BIT SAMPLES/LINE

B

3.1 PERCENT TOLERANCE
230,000 – 9 BIT SAMPLES
213 SAMPLES/LINE
21.3 PERCENT OF TOTAL SAMPLES
S_r = 4.70:1
B_r = 2.61:1

C

6.25 PERCENT TOLERANCE
123,000 – 9 BIT SAMPLES
114 SAMPLES/LINE
11.4 PERCENT OF TOTAL SAMPLES
S_r = 8.78:1
B_r = 4.88:1

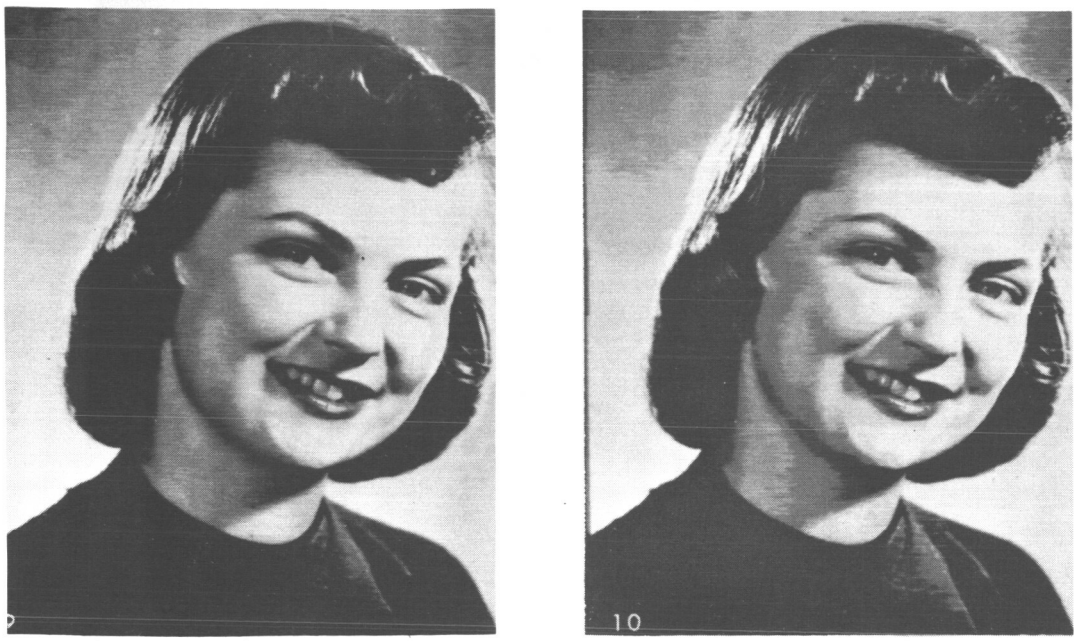
D

9.4 PERCENT TOLERANCE
104,000 – 9 BIT SAMPLES
94 SAMPLES/LINE
9.4 PERCENT OF TOTAL SAMPLES
S_r = 10.59:1
B_r = 5.33:1

Figure 12-7. Reduction By Step Method

IMPORTANCE OF PROPER RECONSTRUCTION

The false contouring referred to above results from large changes in the data between nonredundant samples. The fan method does much to break this up by simply connecting the nonredundant samples with a straight line rather than a step. Figure 12-9 demonstrate the extreme importance of the accuracy to which the line is reconstructed. Both pictures were coded to six bits, the redundancy removed, and then only the four most significant bits were transmitted. Figure 12-9(b) reconstructs the straight line between nonredundant samples to only four bits while 12-9(a) reconstructs to eight bits. This is possible because of the operation of the fan method, which is a patented process of Radiation Inc. Although the two nonredundant samples are known only to 4 bits, when the division is performed to replace the redundant samples along the straight line, it is carried out to 12 binary bits and reconstruction is performed to the 8 most significant bits. The more bits to which this division is carried out, the better the



FAN METHOD – 3.9 PERCENT TOLERANCE
132,000 – 4 BIT SAMPLES
RECONSTRUCTED TO 8 BITS

FAN METHOD – 3.9 PERCENT TOLERANCE
135,000 – 4 BIT SAMPLES
RECONSTRUCTED TO 4 BITS

Figure 12-9. Control of False Contouring

reconstructed samples will fit the straight line. The only difference between Figures 12-9(a) and 12-9(b) is that only four most significant bits of the division were used for Figure 12-9(b), while eight were used for Figure 12-9(a).

A further improvement of this effect could be made by operating the reconstruction process at a higher sampling rate than was used for the reduction process. In other words, there would be more samples along the straight line, and, when filtered by an interpolation filter, the analog output waveform would more nearly match the input waveform.

RESULTS OF HYBRID METHOD

When a sampled waveform is reconstructed by interpolation with a higher order process, it will produce fewer errors, or, if the same error is desired, the original sampling rate can be reduced. By the same token, a waveform can have the redundancy removed with the zero-order process and be reconstructed with a linear interpolation process. This is known as a hybrid redundancy reduction process. Its main advantage is that the reduction equipment can be kept simple. It does, however, have several problems that must be considered. In certain waveforms, the peak error produced in reconstruction may be twice the tolerance; as, for instance, when a waveform almost goes out of tolerance in a positive direction and then suddenly goes out of tolerance in a negative direction. If a straight line is drawn from the original sample to the sample that is just out of tolerance, at the peak which almost went out of tolerance there will be a peak error greater than the tolerance. In the TV case where sharp edges appear, they may be fuzzy when reconstructed by this method.

When a four-bit timing word is used, however, neither of these problems is likely to cause trouble. The chance of the first problem occurring during a maximum of 16 samples is extremely small while the amount that a sharp edge will be shifted by this method will be only a small part of the 16 samples maximum, which cannot be seen by the eye.

In addition to the simplicity of the reduction process, the linear interpolation process requires a lower sampling rate than a zero-order or step process.

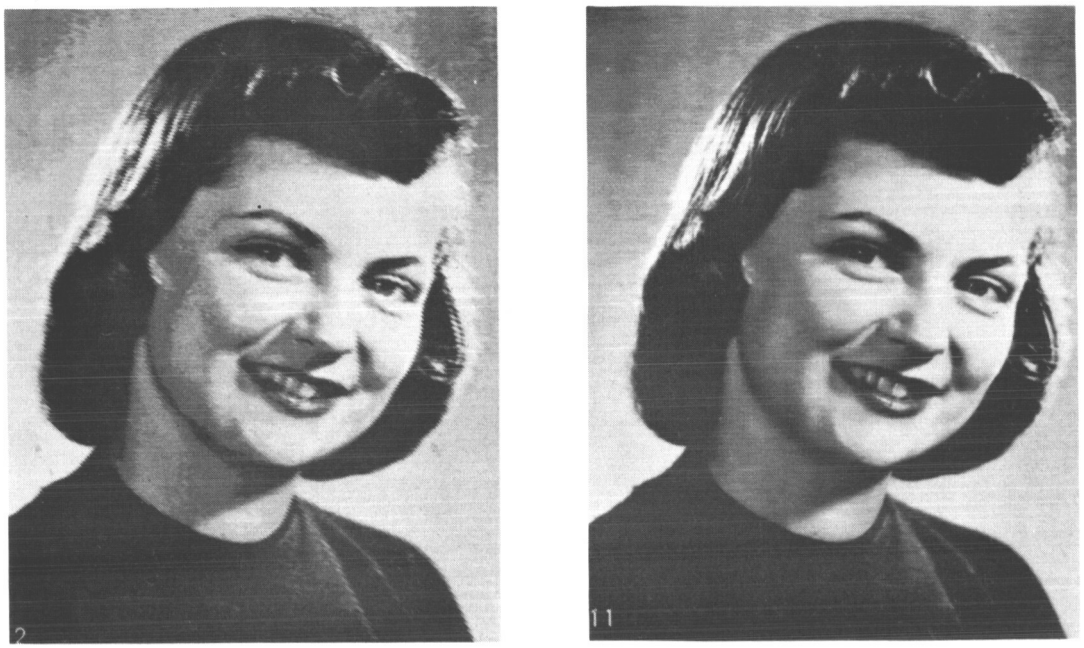
The main advantage of the hybrid method over a straight step method, when used on TV, is that the large steps that produce false contouring are broken up. This can be seen by examining Figure 12-10(a). The top half of the picture was reduced and reconstructed by the step method while the bottom half was reconstructed by a linear interpolation. This is, however, a four-bit picture, which produces very bad false contouring. When the same process is carried out on a picture of five or more bits, the difference is not very much. Figure 12-10(b)

Figure 12-10. Operation of Hybrid Redundancy Reduction Method

shows the result of a straight step reduction on a five-bit picture. This again shows the importance of always transmitting a five-bit picture, and that the hybrid method is not needed when so doing.

UNIFORM SAMPLING COMPARED WITH REDUNDANT SAMPLES

As mentioned earlier, any redundancy reduction scheme merely selects the minimum number of samples and interpolates between them. It is extremely important that there is a sample at the right location on the waveform to allow reconstruction to the desired accuracy. The importance of this statement can be seen by examining Figure 12-11. Figure 12-11(a) has been sampled at a uniform rate of 189 four-bit samples per line, and reproduced with no further processing. Figure 12-11(b) was sampled 1000 times per line using four-bit samples. The redundancy was removed by the fan method at a tolerance of 6.25 percent or 1 part in 16. This reduced the samples to 94 eight-bit samples



UNIFORM SAMPLING
204,000 – 4 BIT SAMPLES
816,000 TOTAL BITS

REDUCED FROM CORRECT SAMPLING RATE
102 – 8 BIT SAMPLES
816,000 TOTAL BITS

Figure 12-11. Importance of Correctly Locating Samples On Waveform

per line. The two pictures, therefore, have the same total number of bits — 816,000. Even though the picture on the left has twice as many samples as the one on the right, the samples do not fall at the correct location and thus produce a combination of ragged lines and false contouring.

CONCLUDING REMARKS

The importance of recognizing the basic information content of the data to be sampled cannot be overstressed. As has been noted, there has been a complete disregard of the problem throughout industry. Although the transmission of a waveform is not the final answer to data compression, it is an important first step as most telemetry systems in use today rely on it.

It is also important to note that the real criterion for judging the results of redundancy reduction processes on television pictures is not compression ratio,

but is the minimum number of bits necessary to transmit a picture of a given quality.

The fact that two different problems exist in transmitting pictures should not be overlooked. One occurs when it is important to reduce power and not necessarily bandwidth, and the other, where it is important to reduce bandwidth and not necessarily power. The first is the most likely in the case of a satellite, and the second, for transmission over existing ground circuits. When bandwidth is not important, it is possible to operate without a buffer by using an amplitude modulation scheme, thus leaving holes in the data. Location of transmitted words could be determined by flywheel synchronization on the ground. Thus, the timing word could be dropped and reduction greatly improved, and all problems of buffer and channel noise in the timing word would be solved. The peak power would be high but the average power would be greatly reduced.

Another important problem studied, but not discussed here, is the effect of channel noise on the reduced data. When the redundancy is removed, the remaining samples are much more vulnerable to channel noise. Much study has been given to this problem, most of it based on error-correcting, band-spreading types of codes. Unless a great deal of band spreading is done and the length of the corrected words are long, the gains are small. The most satisfactory method appears to be so to arrange the data that when a mistake is made, it will be only one level, which will not be very objectionable. Errors in timing are also disastrous; however, these can be corrected to within reason by using a real-time code and correcting it when noise causes it to fall out of sequence.

Considerations have been given to the design of an airborne system to accomplish what has been demonstrated here. It appears that the system could be built, using all integrated circuits, in a box 4 by 4-1/2 by 9 inches or 162 cubic inches; this would include a 5000-word buffer. It would weigh about 12 pounds and require 7 to 10 watts. It would operate up to 100,000 eight-bit samples per second, which would be equivalent to a general purpose computer with a cycle time of 0.1 microsecond. This would include the analog-to-digital converter but not the scanning system.

The results shown here are believed to be quite pessimistic. It is believed sure that with careful design of the scanning system and attention to the conversion noise problem, the amount of reduction and quality of the pictures can be improved.

EFFECTS OF FALSE CONTOURING

False contouring is an effect that results from the combination of large quantizing levels and the eye's ability to detect very small changes in intensity when there are large areas to compare. A large area of slowly varying intensity, such as a sky background, is quite common in most pictures. If, for example, a density wedge is scanned with a one-bit sample, it will produce two gray levels. As the continuous wedge is scanned, there will be a place in the picture where the quantization level change will occur in approximately the same place and, in turn, produce a false line or contour down the middle of the picture. When the sample is increased to two bits, the four quantizing levels will produce three lines and four different gray levels. As the number of quantizing levels is increased, the number of false lines increases and they become closer and closer together. With a five-bit picture or 32 quantizing levels, the lines are close enough together that the eye does not notice them.

COMPARISON OF FOUR-BIT AND FIVE-BIT REDUCED PICTURES

Although the false contouring of a digital four-bit picture is not particularly bad when it has been reduced by any redundancy reduction process the effects become much more noticeable. This can be seen in Figure 12-8. Figure 12-8(a) has been reduced from a 4-kc, 8-bit picture by using the fan method with 6.25 percent tolerance. Only the five most significant bits of the 98,000 nonredundant samples were transmitted and used for reconstruction. Figure 12-8(b) has been reduced to 107,000 nonredundant samples with the same tolerance where only the four most significant bits were transmitted. This comparison shows graphically the need for transmitting a five-bit picture. Increasing beyond five-bits does not, however, improve the picture significantly.

When the step process is used with large tolerances the results are even more disastrous. This can be seen in Figures 12-8(c) and 12-8(d). Figure 12-8(c) is a five-bit picture reduced to 102,000 nonredundant samples. Note the false contouring appearing on the left side of the face and neck. When a four-bit picture is reduced as in Figure 12-8(d), the effect is extremely bad. If a five-bit picture were reduced to 57,000 nonredundant samples, using the step method and five-bit timing word as was done for the fan method, the picture would probably be worse than 12-8(d).

The effect shown is not one of conventional false contouring but a new one produced by the large value changes between nonredundant samples. It can be controlled by proper reconstruction methods.

A

FAN METHOD
6.25 PERCENT TOLERANCE
5 BIT PICTURE
819 BITS/LINE
98,000 SAMPLES
S_r = 11.02:1
B_r = 6.12:1

B

FAN METHOD
6.25 PERCENT TOLERANCE
4 BIT PICTURE
792 BITS/LINE
107,000 SAMPLES
S_r = 10.09:1
B_r = 5.01:1

C

STEP METHOD
9.4 PERCENT TOLERANCE
5 BIT PICTURE
850 BITS/LINE
102,000 SAMPLES
S_r = 10.59:1
B_r = 5.88:1

D

STEP METHOD
9.4 PERCENT TOLERANCE
4 BIT PICTURE
752 BITS/LINE
102,000 SAMPLES
S_r = 10.59:1
B_r = 5.33:1

Figure 12-8. Comparison of Five and Four Bit Reduced Pictures

APPENDIX A
ABSTRACTS OF UNAVAILABLE PAPERS

Contributors	I. Eisenberger JPL	W. Higgins Sylvania
	E. Posner JPL	J. Purcell GSFC
	D. Bourke JPL	R. Muller GSFC
	D. Slaughter JPL	J. Metzner NYU
	R. Trost JPL	S. Golomb USC
	E. Baghdady ADCOM, Inc.	J. Massey Notre Dame
	J. Schwartz IDA	S. Schwartz Michigan
	R. Barker Yale	D. Weber Lockheed
	D. Schaefer GSFC	C. Kortman Lockheed

13. USE OF SAMPLE QUANTILES FOR DATA COMPRESSION OF SPACE TELEMETRY*

I. Eisenberger

Jet Propulsion Laboratory, CIT
Pasadena, California

On the assumption of normal parent populations and a large sample size the following uses of sample quantiles for data compression are discussed: (1) Estimating the mean and standard deviation, (2) two goodness-of-fit tests, (3) testing the mean of the population, (4) testing the standard deviation, (5) testing the mean and standard deviation simultaneously, (6) two-sample tests, and (7) tests of independence and estimation of the correlation coefficient.

*This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS 7-100, sponsored by the National Aeronautics and Space Administration.

14. EPSILON – DELTA ENTROPY AND DATA COMPRESSION

E. C. Posner

Jet Propulsion Laboratory, CIT
Pasadena, California

A unified framework for the theory of data compression is presented. In the theory, one starts with a data source which is abstracted into the notion of a probabilistic metric space (PMS). Briefly, a PMS is a metric space together with a probability distribution subject to certain mild requirements. The metric on the space corresponds to a "fidelity criterion;" that is, the distance between two points (between two experimental outcomes) is a measure of the loss of fidelity if one outcome occurs but the second is assumed to occur.

The probability distribution on the PMS describes the probability law governing the experimental outcomes. The problem of data compression is not to observe an outcome at a remote point, but to transmit only certain information about the outcome. The transmitted information can be regarded as an indication of which set of possible outcomes occurred.

Thus an epsilon-delta partition of a PMS for positive values of ϵ and δ is defined as a covering of all or part of the space by disjoint measurable sets having a diameter which is at most epsilon. The part of the space not covered has a probability of at most delta. The epsilon represents the fidelity required and the delta represents the allowable failure probability. For example, an A-D converter has a delta representing the offscale probability.

The data compression system works in the following manner. An outcome is observed, and the set of the epsilon-delta partition into which the outcome falls is transmitted. The epsilon represents the maximum uncertainty about the outcome when it is known into which set the outcome falls. It is felt that this concept covers all imaginable data compression techniques.

To define the epsilon-delta entropy of a PMS, one must first define the entropy of an epsilon-delta partition. This entropy is merely Shannon's entropy of the discrete probability space obtained from the partition by regarding each set as a point of the same probability as the set when the probabilities are normalized so as to add up to unity.

The epsilon-delta entropy of the PMS is now defined as the infinitum of the entropies of all epsilon-delta partitions of the PMS. This entropy is essentially the minimum number of bits necessary to describe outcomes with a precision of at least epsilon, when, at most, delta of the outcomes do not have to be handled by the data compression system.

The problems of data compression for a given data source are described as follows: First, create a PMS which is a good model of the probability law and the fidelity law of the experiment. Second, find out which epsilon is needed for the required precision. Third, accept a certain failure probability delta. Fourth, find the epsilon-delta entropy, or a good approximation to it. Fifth, settle on a certain efficiency that you are willing to accept. (For example, an efficiency of one-half means that twice as many bits are being sent as the absolute minimum given by the epsilon-delta entropy. The efficiency concept is regarded as a better choice than a "data compression ratio" concept.) Sixth, find an epsilon-delta partition of the PMS which achieves the required efficiency in a reasonably mechanizable fashion.

This mechanization amounts to finding which partition set an outcome belongs in. This could be the most difficult part of data compression. For example, if one has a PMS of functions on an interval, the first 20 Fourier coefficients might be sufficient to compute and transmit if the distance is given by a square error criterion.

The model presented covers almost any conceivable data transmission problem. A mathematical difficulty, however, is that of finding epsilon-delta entropies. There are a few results for the case in which the PMS is the space of square integrable functions on a closed and bounded interval, with probability induced by a Gaussian stochastic process on the interval. This model, however, covers a wide range of data sources which produce functions of time.

In closing, it should be pointed out that the theory of PMS's and their epsilon-delta entropies is still under very active development. This work is the result of a joint effort of the author and Howard Rumsey, Jr., also of JPL.

15. DATA COMPRESSION AND DATA USERS

D. G. Bourke*

Jet Propulsion Laboratory, CIT
Pasadena, California

Data compression will achieve a reasonable decoupling of the data sources from the limitations of the communications channel, but it necessitates much closer liaison with the data users. Some of the design constraints and requirements imposed by the data users during the conceptual and preliminary design of an engineering data system which incorporated data compression are discussed. The continuous interaction with the data users was a great aid in solving the attendant design problems.

As a secondary benefit derived from intimate data user liaison, much valuable groundwork was laid for later phases of the system development.

*Mr. Bourke is now an employee of IBM, Federal Systems Division.

16. ADAPTIVE DATA CONTROL AND PROCESSING FOR SCIENTIFIC EXPERIMENTS WITH ON-BOARD PLANETARY PROBES

D. W. Slaughter

Jet Propulsion Laboratory, CIT
Pasadena, California

Criteria are presented which influence the design approach to adaptive systems for the acquisition, encoding, processing, and management of scientific data onboard interplanetary spacecraft. Emphasis is placed upon system considerations rather than upon hardware cost limitations, although practicality is not ignored. The criteria considered include the scientific objectives of the mission, characteristics of the sensors and ancillaries, and the utilization of ground commands.

The techniques which the scientist uses to process and evaluate the received flight data, including autocorrelation of the data from several sensors, will be evaluated for their effect on the design of onboard systems. The possibility for unexpectedness in the scientific phenomena (or deviation from the scientist's model) will be considered; also covered are the scientist's anxiety concerning the masking or rejection of unexpected data by the onboard systems and the generation of false messages, including those generated by failure modes in the adaptive system.

17. A PRIORI AND A POSTERIORI STUDIES OF ADAPTIVE TELEMETRY TECHNIQUES FOR DEEP SPACE MISSIONS

R. F. Trost

Jet Propulsion Laboratory, CIT
Pasadena, California

During the past several years the Spacecraft Telemetry and Command Section at the Jet Propulsion Laboratory (JPL) has been studying two very general classes of spacecraft data, namely, engineering telemetry data and scientific video data. Although some a priori studies were performed in each class, most studies were done in an a posteriori manner.

An important result from the a posteriori studies of engineering telemetry data was the development of a spacecraft Engineering Data Handling System (EDHS). Direct consultations with the data users dictated the system organization and characteristics of the EDHS.

Basically, the EDHS is divided into four subsystems. Each has a transfer function which is considered best for the subset of measurements which it processes. These subsets are termed operational, operational-performance, and performance. They are processed by the F1 Absolute Rate Commutator subsystem, the F2 Data Compression (with a minimum time-delay feature) subsystem, and the F3 Data Compression subsystem, respectively.

The characteristics of the F1 subsystem are quite simple. Measurements are sampled at one of two constant rates regardless of changes in the communication link bit rate. The constant rates are stored in a small memory and selected individually via a "user control line" which is sampled simultaneously with the user's measurement. Operational measurements processed by the F1 subsystem are usually those which are highly critical to the success of the mission.

The characteristics of the F2 and F3 subsystems are very similar, with one exception. Although both contain data compression and the attendant buffer memories, the F2 subsystem also has a minimum time-delay feature which permits some of the more important data to bypass the buffer and be received in real time. It is necessary because many of the operational-performance measurements processed by F2 are normally very static; however, sudden

abnormal changes must be relayed to the receiving station as soon as possible if corrective measures are to be initiated.

In addition to the F1, F2, and F3, there is a fourth subsystem called the Master Multiplexer. This is simply a time multiplexer with a programmed priority scheme for accepting the outputs from the other three subsystems and producing a continuous data bit stream to the modulator.

Finally, it should be mentioned that a command reprogramming capability exists for updating processing parameters for each channel. Also, a provision has been made for periodic confidence sampling which produces a complete readout of all channels at predetermined times.

In addition to the engineering data studies, an a posteriori video data compression study is also currently in progress at JPL. Television pictures from the Ranger and Mariner spacecrafts are being compressed by the use of various algorithms for compression and addressing. The compression and reconstruction is done with a digital computer, and preliminary results of data to which no noise has been added is favorable.

For the selected Ranger IX picture, a typical Net Compression Ratio (NCR), which includes addressing information, of about unity was realized using a first-order predictor with an aperture of K = 1 (i.e., it is not information destroying). For the same aperture, the first-order interpolator yielded an NCR of 1.2. When the aperture is increased (K = 2), the first-order predictor yielded an NCR of 1.4, and the first-order interpolator yielded an NCR of 1.8.

A typical NCR range for the first eighteen Mariner IV pictures was 1.3 to 2.4 when they were compressed using the first-order predictor with K = 1. For the same aperture, the first-order interpolator yielded an NCR range of 2.5 to 9.0. When the aperture was increased (K = 2) the respective NCR's were also increased by a factor of two or three.

These topics have been presented to permit others to understand better the type of work currently in progress in adaptive telemetry within the Spacecraft Telemetry and Command Section of the Telecommunications Division at the Jet Propulsion Laboratory.

18. ADAPTIVE MODULATION AND DEMODULATION FOR TELEMETRY

E. J. Baghdady

ADCOM, Inc.
Cambridge, Massachusetts

The signal design objectives for the most efficient utilization of available primary power on board a spacecraft are outlined. Methods for implementing the necessary steps are described, evaluated, and compared.

For spacecraft/ground communication links having a capacity which varies over the course of a mission, some form of data rate adaptation may be desirable. This paper considers the problem of preselecting a set of discrete data rates optimized under various criteria for the expected channel variation. The role of feedback communication (repeat-request) as a data rate adaptation technique is also discussed. Additionally, a feedback demodulation technique for the adaptive demodulation of FM/FM telemetry is described and evaluated.

19. BIT-PLANE ENCODING

J. W. Schwartz

Institute for Defense Analysis
Arlington, Virginia

and

R. C. Barker

Yale University
New Haven, Connecticut

Bit-plane encoding, a source encoding technique designed for use with data gathered in space probes, is described. The application of the technique to certain data gathered by the Explorer XII satellite is discussed. In addition, the general approach to adaptive telemetry taken in the course of the research is presented.

Bit-plane encoding is intended for use aboard spacecraft.*† The encoder implementation consists of a memory to store data samples, a monitor, and a code box. Sample values in binary form are received sequentially and stored in the memory. While the values are being stored, the monitor makes certain measurements on each bit plane to determine how each plane is to be treated. When a complete group of values has been stored, the encoding procedure begins. First some bits are read out of the monitor to indicate the operation which the code box will perform on each bit plane; then the monitor controls the readout of the memory, one bit plane at a time, and selects the operations to be performed by the code box. Both the monitor and the code box perform simple operations on binary sequences.

*Schwartz, J. W., "Data Processing in Scientific Space Probes," Tech. Note 6, Yale Univ., September 1963.

†Schwartz, J. W., and Barker, R. C., "Bit-Plane Encoding: A Technique for Source Encoding," IEEE Trans. on Aerospace and Electronic Systems, AES-2(4):385-392, July 1966.

Bit-plane encoding is especially useful when the data have an amplitude spectrum which is concentrated in different ranges in different time intervals. With a stored group of 128 samples, bit-plane encoding could be used to describe energetic particle counts gathered by Explorer XII with less than 50 percent as many bits as were actually used and with no loss of information. The technique also conveniently allows certain useful information-destroying operations.

20. EXPERIENCES WITH SMALL ADAPTIVE DATA PROCESSORS

D. H. Schaefer

Goddard Space Flight Center
Greenbelt, Maryland

Data processors have flown in Goddard's small scientific satellites starting with Vanguard III. Two of these processors with adaptive attributes are described.

The first of these devices is an optical aspect data processor that flew in the Atmosphere Explorer satellite. This processor had effectively nine separate digital inputs from Sun, Moon, and Earth sensors. It was desired to determine which of these had had an input during a frame of telemetry; and, furthermore, in which sixteenth of the frame period certain of these inputs had arrived. Only eight bits of telemetry per frame were allocated for the transmission of all this information. By use of a variable format where the first bit sent determined the meaning of all subsequent bits, and by the use of a four-level priority system, enough information was transmitted to determine the aspect of the satellite.

The second device described is an experimental digital tracking device that has been designed to "clean up" the noisy signals from rubidium vapor magnetometers so that onboard processing can be accomplished.

This device operates by sampling the magnetometer signal and its noise. The signal is sampled at times determined by a voltage-controlled oscillator. To begin with, the device samples the input signal every half cycle of the voltage-controlled oscillator. If at the first sample time the sampled voltage is positive, at the next time, negative, at the next time, positive, and so forth, the oscillator is tracking the signal and no corrections are necessary. Violation of this prime rule leads to further tests to determine whether the frequency of the voltage-controlled oscillator is high or low. A given number (say, sixteen) of "high" or "low" decisions in a row must be made before any correction pulses are given to change the frequency of the voltage-controlled oscillator so that noise will not initiate incorrect commands.

The device can also be used to determine if a signal is present in the noise, and, if not, to switch the phase of the feedback to the magnetometer so that a signal can be produced. The decision that no signal is present is made when an approximately equal number of "high" and "low" correction indications are being

produced. If this is the case, it is assumed that only noise is present and switching of the feedback takes place.

21. A DATA MANAGEMENT SIMULATION EXERCISE

William F. Higgins

Sylvania Electric Product, Inc.
Waltham, Massachusetts

A data management simulation that has been programmed on a general-purpose computer is presented. The simulation is aimed at defining the queue buffer control requirements for adaptive processing. The input data in the simulation model contain priority and nonpriority data, but priority data are not compressed in the simulation. The compressor used in the simulation model for nonpriority channels is a zero-order floating aperture predictor; the aperture is a function of system control parameters. The control is aimed at compressing or deleting nonpriority data in such a way as to maintain a reasonable amount of data precision and to make efficient use of the queue buffer.

The input data are contained in 26 channels. Priority data comprise 10 percent of the data flow, and nonpriority, 90 percent. The difference between the input data flow and the output data flow per unit time is $\dot{q}$; this difference is accumulated in a buffer. The number of data words in the buffer is called the queue, q. The queue buffer control equation is a function of q and $\dot{q}$. The buffer control has been aimed at preventing the queue from underflowing or overflowing.

The input data are Gaussion. The initial simulation runs using various control functions were generally unsatisfactory. The buffer was used inefficiently and control was maintained at the price of data precision. The input-output rate of the system is an important parameter in the control function. For a particular input-output rate, the aperture on the controlled (nonpriority) channels for steady-state operation can be expressed in terms of standard deviations of the input data. For an input-output word rate of 10 to 2, the aperture averaged 2.3 standard deviations for a simulation run. This was expected and it validated the program. To maintain precision, it was necessary to modify the control equation by having nonpriority channels deleted in a specified way as a function of the aperture. This improved control was paid for by having nonpriority channels occasionally deleted. The duty cycle for these channels was a function of the input-output rate for a given input sequence.

The nonpriority channels were modified to have different statistics, and the control function had to be tailored to particular sets of channels. Although the simulation runs were terminated before the nonstationary data control was fully

defined, certain general observations could be made. When there is enough correlation between channels of data, these may be processed adaptively as a group. However, in general, the control must be tailored to the individual channel. One parameter useful in the control of nonstationary channels is the short-time averaged standard deviation $\sigma_T(t)$.

The simulation exercise, in general, has emphasized the need to tailor the processing and control to the experiments and/or sensors involved.

22. ADAPTIVE TELEMETRY – PAST PERFORMANCE AND FUTURE POTENTIAL

J. Purcell and R. Muller

Goddard Space Flight Center
Greenbelt, Maryland

The telemetry systems that were designed to meet the requirements of large scientific satellites are examined. These spacecraft carried many experiments with varied output data characteristics, and thereby provide good examples for missions where all experimental data cannot be processed by some general rule. The benefits and problems created by the adaptability in these systems are used as a basis to forecast the impact of increased adaptability on future spacecraft programs. Examples of onboard data processors now under consideration are presented.

23. SIGNAL POWER CONTROL ACCORDING TO MESSAGE INFORMATION CONTENT

J. J. Metzner

New York University
Bronx, New York

The level of message information contained in satellite telemetry data can vary widely with respect to time. It can be highly inefficient to provide at all times for the maximum rate of data transmission. Compressive message encoding, a method which improves efficiency, requires substantial buffering storage to minimize lost data during the periods of high message information content.

In many space communication applications, the total energy transmitted is likely to be far more of a limitation than instantaneous peak power. Methods by which transmitted power and the transmitted data rate can be varied according to the needs dictated by the message information content are proposed. Such a distribution of available energy constitutes an efficient utilization of communication resources. Specifically, it can serve either of two purposes: In conjunction with compressive coding, it can serve as a replacement for the bulk of the buffer storage requirements, as well as improve communication efficiency, or, if certain transmitting procedures prove to be practical, it can replace compressive coding itself as a means of improving efficiency.

The efficient use of energy is discussed, and it is pointed out that, where the bandwidth is of little or no consideration, the total information that can be transmitted in any time interval is limited only by the total signal energy employed during that time. It does not depend upon the uniform distribution of energy within the time interval. Where bandwidth is limited, there is some loss due to the nonuniform distribution of energy, but the procedure of varying the power retains much of its usefulness.

Several ways for varying the power and data rate are suggested. One involves changing the number of waveforms in time T, providing two or more different modes of operation. A second involves changing the time base; that is, in order to send at twice the rate, the waveforms would be squeezed to occupy one-half the time while the power was doubled. A third scheme involves the use of an auxiliary transmitter only during periods of high information content.

Fourth, a method could be envisioned whereby, if a sample value is the same (within some tolerance) as its previous value, nothing is sent and (ideally) no power is used except for synchronization. When a sample does change, the value of the sample is transmitted with full power. This latter method, which is applicable only in the unconventional situation where total transmitter power utilization in the unmodulated mode is far less than that in the fully modulated mode, has the potential of serving as a replacement for compressive coding.

Other factors of the signal power control are also considered in conjunction with the variation according to message information content. They are: control by command, by preprogramming according to distance, by message priority, and by the current status of the energy reserve.

24. CODING PROBLEMS OF ADAPTIVE TELEMETRY

S. W. Golomb

University of Southern California
Los Angeles, California

There are three completely different approaches to "sophisticated telemetry." The most traditional of these is to seek efficient channel utilization in the sense of Shannon, which involves the consideration of the communications link as a common carrier to be upgraded by matched filters or error correction or other techniques independent of the data being communicated. The second level of penetration into the problem is properly called data compression and assumes an a priori model for the type of data which will probably be encountered. In terms of this model, salient parameters of the data (e.g., histogram levels or quantiles) are computed at the transmitter terminus, and only these compressed data are communicated. The approach with the third degree of ingenuity and sophistication is that of designing the data gathering and preprocessing equipment in such a way that the processing will adapt to the phenomena actually encountered. Only this third-level approach is accurately termed adaptive telemetry, and there are persuasive political-technical reasons why little if any adaptive equipment is likely to be tried in space communications in the near future. Since "efficient channel utilization" is now generally appreciated, the significant research and development emphasis for the next several years should be in the data compression area. The coding problems here involve the extraction, representation, and protection of the significant portions of the experimental data.

25. SCHEME FOR ADAPTIVE THRESHOLD DECODING

J. L. Massey

University of Notre Dame
Notre Dame, Indiana

The additive Gaussian noise channel is a theoretical model which corresponds closely to the actual channel in space communication systems. The usual approach to the problem of reliable data transmission through this channel employs binary antipodal signal segments with successive segments chosen in accordance with the digits of a binary error-correcting code. The maximal-length codes are the class of binary block codes for which the resulting transmitted waveforms generate the simplex structure in signal space. The performance of these codes on the Gaussian channel has been tabulated by Viterbi*.

A class of nonblock binary codes, the convolutional uniform codes found by the author†, give essentially the same performance as the equivalent-rate, maximal-length codes. Moreover, the uniform codes are simpler to encode, an important consideration for space vehicles where there is a premium on encoder simplicity. The interesting feature of the uniform codes is the ease with which the code rate can be changed to match the changing conditions of the signal-to-noise ratio as occurs when the distance of a space probe from the Earth increases.

For any integer m, there is a binary, uniform code with rate $R = 2^{-m}$ and a minimum distance $d = (m + 2)2^{m-1}$. The same encoding circuit for rate $R = 2^{-M}$ can be used for all rates $R = 2^{-m}$, $m = 1, 2, \ldots M$, simply by disabling the last M - m stages of the M-stage generator. The fact that the rate is an inverse power of two facilitates synchronization of the data bits and the 1/R times as numerous encoded bits. (This is in contrast with the maximal-length case where there are diophantine problems associated with rate changing.)

Moreover, the uniform codes can be decoded very simply by a threshold decoder‡ which uses to full advantage the likelihood information for the received

*Viterbi, A., "Phase-Coherent Communication over the Continuous Gaussian Channel," in Digital Communications with Space Applications, S. Golomb, Ed., Prentice-Hall, 1964, pp. 106-134.

†Massey, J., "Uniform Codes," to appear in IEEE Trans. Info. Th., IT-12, April, 1966.

‡Massey, J., "Threshold Decoding," M. I. T. Press, 1963.

bits. This permits real-time operation with simple decoding equipment and essentially maximum likelihood decoding performance. The decoder can be adapted for rate changes in a manner very similar to that for the encoder. It can also be shown that, unlike some forms of nonblock decoding, there is no danger that a decoding mistake will trigger a long succession of further decoding mistakes. This can be important in applications such as the space channel where a feedback link to the encoder is not readily available.

26. AN EMPIRICAL BAYES TECHNIQUE IN DETECTION AND ESTIMATION THEORY

S. Schwartz

University of Michigan
Arbor, Michigan

An empirical Bayes technique is applied to the class of problems in which a sequence of information-bearing signals is assumed to be a stationary, random process with the underlying probability structure unknown. The observations are added to Gaussian noise with a known, nonwhite spectrum. The statistical problem is to extract information from each member of the sequence by estimation or detection, depending on the problem. The Bayes technique involves the use of accumulated past observations to obtain consistent estimates of the unknown distributions or related quantities. These are used to form a sequence of one-stage decision (estimation) procedures which converge to the optimum procedure one would use if all pertinent distributions were known.

This empirical technique differs from other learning procedures and from what has been called nonsupervised pattern recognition in a number of respects. An important difference is that a statistical dependence on the information-bearing signals and on the additive noise is permitted. Second, the signals (at the receiver) are taken as random processes as opposed to fixed (but unknown) signals buried in the noise. Third, the method of estimating the unknown distributions is an improvement over the other methods suggested in the past in that it gives faster convergence and simpler implementation.

27. ADAPTIVE DATA COMPRESSION TECHNIQUES

D. R. Weber and C. M. Kortman

Lockheed Missile and Space Co.
Sunnyvale, California

Adaptive telemetry is a term that has gained considerable popularity over the past several years. To assist in the definition of this term it is desirable to relate adaptivity to three levels of sophistication. The first and simplest form of adaptivity is the "adaptable" system. An adaptable process is one which reacts in a rigid manner to one or more outside parameters or commands. The second is the "adaptive" system, and is defined as a process which reacts to parameters monitored within itself. The third and most sophisticated form of adaptivity is the "self-adaptive" system. A self-adaptive process involves monitoring, learning, and reaction. In a truly self-adaptive system the reaction process is invented as an outcome of the learning process. The majority of techniques presently used in data compression fall within the areas of adaptable or adaptive processes. Four basic categories of data compression, i.e., parameter extraction, adaptive sampling, redundancy reduction, and encoding are recognized. The remainder of the paper is devoted to the application of redundancy reduction to video data.

In measuring the effectiveness of several different redundancy reduction algorithms, it is necessary to present bandwidth compression ratio as some function of error. Attempts to use peak and rms errors were not successful since there was little correlation between these measures of error and the subjective evaluation of the experimenters. Hence, it was necessary to relate bandwidth compression ratio to the subjective analysis of picture fidelity. The results of a study sponsored by the Goddard Space Flight Center showed that Tiros satellite pictures can be compressed successfully, resulting in bandwidth compression ratios of four to one. Compression ratios of five to one were achieved on Ranger pictures, and six to one on scanned Gemini photographs. The first-order polynomial interpolators were shown to be more effective than predictors and the zero-order interpolator.

The use of an adaptive aperture and adaptive filter techniques as a means of buffer control is presented in detail. Both techniques were successful in minimizing buffer overflow and buffer underflow. However, the adaptive filter control technique resulted in greater picture degradation than adaptive aperture. It was concluded that the high frequency components in pictorial data are of greater

importance for maintaining picture fidelity than are the absolute intensity values. The use of an adaptive aperture based upon the measure of recent redundant sequences proved to be of value for improving overall picture fidelity.

The subject of sampling relative to redundancy reduction was discussed in detail during the question and answer period. In summary, sampling and redundancy reduction should always be considered jointly if the performance of a system is to be maximized. In an undersampled system excessive errors will result from aliasing. If data are oversampled the overall bandwidth requirements will increase since a redundancy reduction process cannot eliminate all redundancies and efficiency will be lost because of the encoding process.

APPENDIX B
LIST OF ATTENDEES

William L. Alford
GSFC, Code 563
Greenbelt, Md.

Prof. V. R. Algazi
Department of Elec. Eng.
Univ. of California
Los Angeles, Calif.

Raymond Fallen, Jr.
GSFC
Greenbelt, Md.

Carroll P. Ancheus
T. J. Watson Research Center
International Business Machines
Bethesda, Md.

Tage O. Anderson
Jet Propulsion Laboratory
Pasadena, Calif.

Dr. Elie Baghdady
ADCOM, Inc.
Cambridge, Mass.

Prof. Richard C. Barker
Yale University
New Haven, Conn.

William P. Barnes
GSFC, Code 563
Greenbelt, Md.

Earl Beard
GSFC, Code 564
Greenbelt, Md.

Henry M. Beisner
T. J. Watson Research Center
International Business Machines
Bethesda, Md.

Lewis Billing
MITRE Corp.
Bedford, Mass.

Dr. Victor W. Bolie
Advanced Technology Corp.
Costa Mesa, Calif.

Joseph B. Bourne
GSFC, Code 564
Greenbelt, Md.

Dinall G. Bourke
T. J. Watson Research Center
International Business Machines
Bethesda, Md.

Carlo J. Broglio
GSFC, Code 226
Greenbelt, Md.

Stanley H. Brown
GSFC, Code 226
Greenbelt, Md.

Thomas E. Burke
NASA Electronic Research Center
Cambridge, Mass.

Willis S. Campbell
GSFC, Code 562
Greenbelt, Md.

A. George Carlton
Applied Physics Lab.
John Hopkins University
Silver Spring, Md.

John B. Carraway
GSFC, Code 562
Greenbelt, Md.

Robert T. Caughey
GSFC, Code 565
Greenbelt, Md.

Frank Charles
Jet Propulsion Laboratory
Pasadena, Calif.

Prof. Yaohan Chu
University of Maryland
College Park, Md.

Rodger A. Cliff
GSFC, Code 711
Greenbelt, Md.

John J. Concordia
Sylvania Elec. Products
Waltham, Mass.

Cyrus J. Creveling
GSFC, Code 561
Greenbelt, Md.

John M. Davies
T. J. Watson Research Center
International Business Machines
Bethesda, Md.

Walter D. Davis
GSFC, Code 521
Greenbelt, Md.

Prof. Lee D. Davisson
School of Eng. & Applied Science
Princeton University
Princeton, N. J.

Dr. U. D. Desai
GSFC, Code 611
Greenbelt, Md.

James O. Duffy
Jet Propulsion Laboratory
Pasadena, Calif.

John L. East
NASA Headquarters
Washington, D. C.

W. E. Edwards
National Security Agency
Ft. Geo. Meade, Md.

Dr. Isadore Eisenberger
Jet Propulsion Laboratory
Pasadena, Calif.

F. H. Emens
MSFC/NASA
Huntsville, Ala.

Leonard Farkas
Institute for Defense Analysis
Arlington, Va.

John F. A. Fisher
National Security Agency
Ft. Geo. Meade, Md.

Prof. Gerard A. Gallagher
Massachusetts Institute of Technology
Cambridge, Mass.

L. W. Gardenhire
Radiation, Inc.
Melbourne, Fla.

J. M. Glass
Sylvania Elec. Products
Waltham, Mass.

Thomas Goblick
Lincoln Labs
Massachusetts Institute of Technology
Cambridge, Mass.

Prof. Solomon W. Golomb
University of Southern California
Los Angeles, Calif.

T. Patrick Gorman
GSFC, Code 565
Greenbelt, Md.

Eugene I. Grunby
GSFC, Code 565
Greenbelt, Md.

E. J. Habib
GSFC, Code 560
Greenbelt, Md.

A. L. Hedrich
GSFC, Code 733
Greenbelt, Md.

Paul Heffner
GSFC, Code 563
Greenbelt, Md.

Nathan Heifetz
GSFC, Code 564
Greenbelt, Md.

Erich Herz
General Dynamics Systems
San Diego, Calif.

W. H. Higgins
Electronic Systems
Sylvania Elec. Products
Waltham, Mass.

Daniel Hochman
Lockheed Missile Space Center
Sunnyvale, Calif.

Gerald J. Hogg
GSFC, Code 521
Greenbelt, Md.

Roger C. Hollenbaugh
GSFC, Code 731
Greenbelt, Md.

Richard G. Holmes
GSFC, Code 563
Greenbelt, Md.

Stephen G. Howie
Martin-Marietta
Baltimore, Md.

R. T. Hynes
NASA Headquarters
Washington, D. C.

Donald E. Jamison
GSFC, Code 564
Greenbelt, Md.

O. L. Jones
GSFC, Code 565
Greenbelt, Md.

Robert E. Jones
GSFC, Code 563
Greenbelt, Md.

Thomas J. Karras
GSFC, Code 563
Greenbelt, Md.

Prof. R. B. Kelley
University of Rhode Island
Kingston, R. I.

C. R. Laughlin
GSFC, Code 731
Greenbelt, Md.

Ira L. Lewis
GSFC, Code 564
Greenbelt, Md.

William C. Lindsay
Jet Propulsion Laboratory
Pasadena, Calif.

Dr. G. Ludwig
GSFC, Code 560
Greenbelt, Md.

Warren Lushbaugh
Jet Propulsion Laboratory
Pasadena, Calif.

Thomas J. Lynch
GSFC, Code 520
Greenbelt, Md.

Gennaro Maccarone
Hdqrs. ESD
Hanscom Field
Bedford, Mass.

D. L. Margolis
GSFC, Code 562
Greenbelt, Md.

Michael Mahoney
GSFC, Code 565
Greenbelt, Md.

Benn A. Martin
Jet Propulsion Laboratory
Pasadena, Calif.

Prof. James I. Massey
Department of Elec. Eng.
University of Notre Dame
Notre Dame, Ind.

Marvin S. Maxwell
GSFC, Code 731
Greenbelt, Md.

Prof. J. J. Metznes
New York University
Bronx, N. Y.

John E. Miller
GSFC, Code 733
Greenbelt, Md.

William H. Miller
GSFC, Code 263
Greenbelt, Md.

James C. Morakis
GSFC, Code 520
Greenbelt, Md.

Ronald M. Muller
GSFC, Code 562
Greenbelt, Md.

B. G. Narrow
GSFC, Code 564
Greenbelt, Md.

C. B. Parker
Department of Defense
Washington, D. C.

Bernard Peavey
GSFC, Code 563
Greenbelt, Md.

Jack D. Pettus
Martin-Marietta Corporation
Baltimore, Md.

Bryan R. Polson
GSFC, Code 563
Greenbelt, Md.

Dr. L. C. Posner
Jet Propulsion Laboratory
Pasadena, Calif.

Prof. William K. Pratt
Depart. of Elec. Eng.
Univ. of Southern California
Los Angeles, Calif.

Joseph Purcell
GSFC, Code 562
Greenbelt, Md.

J. J. Quann
GSFC, Code 565
Greenbelt, Md.

G. L. Raga
Sarasota Division
Electromechanical Research
Sarasota, Fla.

Prof. L. L. Rauch
EE Department
Univ. of Michigan
Ann Arbor, Mich.

L. H. Rhodes
GSFC, Code 563
Greenbelt, Md.

F. C. Roberson
GSFC, Code 565
Greenbelt, Md.

Dr. R. W. Rochelle
GSFC, Code 711
Greenbelt, Md.

A. C. Rosenberg
GSFC, Code 564
Greenbelt, Md.

Prof. D. J. Sakrison
Department of EE
Univ. of California
Los Angeles, Calif.

T. V. Saliga
GSFC, Code 711
Greenbelt, Md.

Dr. Ray W. Sanders
Space General Corporation
El Monte, Calif.

D. H. Schaefer
GSFC, Code 711
Greenbelt, Md.

John H. Schmidt
GSFC, Code 565
Greenbelt, Md.

Jay W. Schwartz
Institute for Defense Analysis
Arlington, Va.

Prof. L. S. Schwartz
School of Engineering & Science
New York University
Bronx, New York

Prof. S. C. Schwartz
University of Michigan
Ann Arbor, Mich.

James E. Scobey, Jr.
GSFC, Code 711
Greenbelt, Md.

Dan S. Serice
NASA Headquarters
Washington, D. C.

Dean W. Slaughter
Jet Propulsion Laboratory
Pasadena, Calif.

J. W. Snively
GSFC, Code 711
Greenbelt, Md.

John Y. Sos
GSFC, Code 563
Greenbelt, Md.

H. R. Stagner
GSFC, Code 565
Greenbelt, Md.

G. R. Stonesifer
GSFC, Code 564
Greenbelt, Md.

F. J. Styles
GSFC, Code 562
Greenbelt, Md.

Dr. C. M. Sussman
ADCOM, Inc.
Cambridge, Mass.

Taro Terashi
GSFC, Code 564
Greenbelt, Md.

Rudolph Trost
Jet Propulsion Laboratory
Pasadena, Calif.

M. Tveitan
Jet Propulsion Laboratory
Pasadena, Calif.

R. L. Van Allen
GSFC, Code 711
Greenbelt, Md.

Prof. C. A. Wagner
Yale University
New Haven, Conn.

William C. Webb
GSFC, Code 563
Greenbelt, Md.

Prof. C. L. Weber
Dept. of Elec. Eng.
Univ. of Southern California
Los Angeles 7, Calif.

D. R. Weber
Advanced Dev'l. Electronics
Lockheed Missile & Space Co.
Sunnyvale, Calif.

Robert W. Wettingfeld
GSFC, Code 514
Greenbelt, Md.

Prof. P. A. Wintz
Purdue University
Lafayette, Inc.

Fred A. Wulff
GSFC, Code 564
Greenbelt, Md.

Howard L. Yudkin
Lincoln Laboratories
Massachusetts Institute of Technology
Cambridge, Mass.